HRSA 94-12

5737

MATHEMATICA
Policy Research, Inc.

Contract No. : 240-94-0052
MPR Reference No. : 8238

COMMUNITY HEALTH CENTER USER AND VISIT STUDY

SURVEY METHODOLOGY REPORT

August 29, 1997

Authors:

George J. Carcagno
Barbara L. Carlson
Rhoda Cohen
Donna Eisenhower

Submitted to:

Bureau of Primary Health Care
Office of Evaluation, Analysis and Research
DHHS, PHS, HRSA
Room 72A2, East-West Towers
4350 East West Highway
Bethesda, MD 20816

Project Officer:
Patricia Royston

Submitted by:

Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ 08543-2393
(609) 799-3 53 5

Project Director:
George J. Carcagno

ACKNOWLEDGMENTS

The authors wish to acknowledge the valuable contributions of project officers Patricia Royston and Jerilyn Regan, and HRSA consultant Joanne Lukomnik. Christopher Moriarity, NCHS statistician, who selected the grantee sample and provided helpful comments on other components of the sample design. A special thank you to the staff of the participating clinics and grantees whose support for and efforts on behalf of the study were critical to its success. The efforts of Sally Waltman, James Cashion, and Anne Self, field coordinators who helped manage the data collection, Patricia Ciaccio, editor, Denise Dunn who was responsible for report production, and MPR's field interview staff are gratefully acknowledged.

CONTENTS *(continued)*

CONTENTS

CONTENTS ***(continued)***

TABLES

EXHIBITS

I. STUDY DESIGN

A. OVERVIEW

1. Background to the Community Health Center Program

From its beginnings in Boston and Mississippi during the 1960s, the community and migrant health center (CHC) program has grown to be a major part of the nation's health care delivery system. CHCs, which at the end of 1995 numbered nearly 600 centers serving about 7 million people, focus on providing health services in areas and to populations that are underserved by the health care delivery system.

Over the years, the program has developed a body of regulations and program expectations that define key features of a CHC. At the same time, however, the program has come to believe strongly in the need to interpret these regulations and expectations at the local level. As a result, each CHC, or "grantee," has taken on unique characteristics that reflect its history and philosophy, and the community it serves. This diversity presents challenges to the design and implementation of a national survey, because the survey organization must be able to adapt its sampling and data collection procedures to accommodate local differences in programs, organizational structures, and records systems.

2. Purpose and Research Questions of the Community Health Center User and Visit Study

The Community Health Center User and Visit Study is a study of CHC medical care users and medical encounters. The study was funded by the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services. The project was under the direction of the Bureau of Primary Health Care (BPHC), the agency responsible for the CHC program, and the Office of Planning, Evaluation, and Legislation (OPEL).

The study was conducted with a nationally representative sample of 48 community health centers. For the user survey, 1,932 randomly selected CHC patients were interviewed about their health status, use of health services, satisfaction with care, access to care, and demographic characteristics. For the visit

survey, a random sample of patient medical encounters was selected and data on health service use during the selected encounters were extracted from 2,878 medical records.' Information from the user survey will be compared with similar information collected in the National Health Interview Survey (NHIS). The NHIS is a continuing household survey conducted by the National Center for Health Statistics (NCHS) with a broad-based national probability sample of the civilian noninstitutionalized population of the United States. Similarly, the information from the survey of CHC medical encounters will be compared with results from the National Hospital Ambulatory Medical Care Survey (NHAMCS), also conducted by NCHS. The NHAMCS obtains data from a nationally representative sample of visits to hospital-based ambulatory outpatient clinics.

The overall purpose of the CHC study was to obtain, for the first time, nationally representative data about the users of CHCs and the services provided to them. These data will provide policymakers with a better understanding of who is being served and the extent to which the CHCs are filling a gap in the health care system. Ideally, this survey would be repeated on a regular basis to provide (1) a database from which profiles of users and encounters can be created, and (2) a basis for monitoring trends over time involving users and services.

The information obtained from the combination of the personal interview survey of CHC users and the separate record-based study of encounters at CHCs will be used to answer a variety of research questions:

- What is the racial and ethnic distribution of CHC users?
- Are certain chronic conditions more prevalent than in the general population?
- Do CHC users need and use more health care than the general population?

'The study design called for a sample of 50 grantees, with 2,000 completed user interviews and 3,000 completed encounter abstractions. Two grantees and their designated replacements refused to participate. The surveys were cut off slightly short of the targets because of schedule and resource constraints.

- What health risk behaviors do these populations present?
- Why do these populations use the health centers? How satisfied are they with the services they receive?
- What are the most prevalent diagnoses?
- Are users with chronic conditions being monitored, and are they following the advice they are given?
- Are age-appropriate preventive services being provided?
- How does the profile of visits to CHCs compare with national studies in conditions, diagnoses, type of provider seen, and disposition?

3. Summary of Survey Methodology

a. User Survey

The survey of CHC users was conducted using a computer-assisted personal interview (CAPI). The survey ranged in length from one-and-a-quarter hours to two hours, depending on the number of relevant questions a particular respondent was required to answer. The survey was designed to be conducted in person at the clinic location identified for the respondent. After several attempts to arrange for an interview, if the respondent agreed to participate but was unable or unwilling to do the interview at the CHC, then the interview was conducted in person at another location (usually the respondent's home) or over the telephone. If the interview was conducted by telephone, a set of response category handcards was sent to the respondent in advance of the interview whenever possible. If the respondent did not have these handcards at the time of the interview, the interviewer read aloud all answer categories which would normally be displayed on the handcards

In summary, there were three possible modes of collection for the user survey:

1. In person at the clinic--while about 80 percent of the interviews were planned to be conducted in this manner, only 38 percent were completed in this setting.
2. Telephone--while about 15 percent were planned to be conducted using this mode, 26 percent were completed by telephone.

3. In person in the home or at another location--while about 5 percent of the interviews were planned to be conducted using this mode, 36 percent were completed in this manner.

The number of user interviews completed for each sampled CHC ranged from 11 to 64. The target number of completes was 2,000 across all CHCs, and 1,932 were completed. More detail on the user survey is provided in Chapters I.B.4 (sample. design); I.C. 1 (CAP1 questionnaire development); II.B. 1 (training); and II.C (implementation including a discussion of response patterns).

b. Visit Survey

This survey was based on a sample of medical encounters at the CHC in 1994. While both user and encounter surveys were conducted at the same sample of CHCs, the two samples were otherwise independent of each other. The encounter survey involved a four-page abstraction form, which was completed either by a CHC staff member or by a trained MPR abstractor. The medical encounter was associated with a medical record, which was pulled to provide the information required to complete the abstraction form. While we originally expected more than 80 percent of the CHCs to prefer to have their own staff collect the encounter data, about half the data were obtained by CHC staff and half by MPR staff.

Between 37 and 88 abstraction forms were completed for each of the CHCs. A target of 3,000 completes was set, and 2,878 completes were obtained. More detail on the visits survey is provided in Chapters LB.5 (sample design); I.C.2 (encounter abstraction form design); II.B.2 (training); and II.D (visits survey implementation including a discussion of response patterns).

B. SAMPLE DESIGN

1. Evolution of the Design

The sample design of the User and Visit Study was broadly planned by HRSA and a consulting statistician from NCHS prior to the issuance of the Request for Proposals. That plan had the NCHS statistician selecting the 50 grantees (the primary sampling unit) and the contractor developing a specific

sampling plan for selecting 2,000 medical users and 6,000 medical visits from these selected grantees and implementing this plan in consultation with the NCHS statistician. While the same selected grantees were to be used for both study components, the user sample and visit sample were to be selected independently; that is, they were not linked to one another.

Four noteworthy modifications occurred to this original plan during the course of the study. All decisions regarding these modifications were made or approved by HRSA based on recommendations from MPR and/or the NCHS statistician. First, in anticipation of the possibility that some of the 50 selected grantees might be unable or unwilling to participate (which, on average, would have meant the loss of 40 users and 120 visit observations), the NCHS statistician added to the sampling plan another stage which would allow for the substitution of a similar grantee if that became necessary. The methodology used for this additional sampling stage is described in more detail below.

Second most grantees had multiple sites, which resulted in a decision to subsample sites. Although ten of the grantees had only one eligible site, the remainder of the grantees had multiple sites, some having more than ten eligible locations--more than had been anticipated.' Dealing with multiple sites creates logistical problems for the interviewers. In many cases, the sites were hours away from one another. To minimize these problems and to conserve resources, a decision was made to subsample sites for the user survey in some circumstances. The decision rules regarding when subsampling would take place and how many would be subsampled were developed after extensive discussion. The final rules for those with more than one eligible site were:

- If the sites were "similar," subsample one site. Whether sites were similar was based on the patient populations served as reported by the grantees (race/ethnicity, percent below 200 percent of the poverty line, urban/rural, general public versus school-based or women-only).

[2]For a site to be eligible, it had to be funded by Section 330 of the Public Health Service Act and provide primary medical services.

- If the sites were "different," take both sites if there were two eligible sites; subsample two sites if there were three or four eligible sites; subsample three sites if there were five or more eligible sites.

For this methodology to be used, certain information was required from the grantees, including site eligibility criteria (type of funding and services provided), number of users and visits within each site, and types of patients served at each site. This information was obtained via three data-collection efforts that preceded the main survey. Obtaining this information in a timely manner proved to be extremely difficult. In addition, the information, once obtained, in some instances proved to be in error and required followup. The methodology used for subsampling sites is described in more detail below, under "Site Sample: Listing and selection."

Third, the definition of the sampling unit for the visit component was modified. Initially, the sampling unit was supposed to be a medical visit, but after determining how CHCs maintain their records and thinking more about how these data would be analyzed, it was decided to change the sampling unit to a medical encounter. The abstraction form and sampling instructions were modified accordingly. "Medical visit" and "medical encounter" are defined below under "Encounter Sample: Definitions and types." It should be noted that, in most circumstances, it makes no difference which unit is used.

Fourth, the targeted number of encounter abstractions was reduced from 6,000 to 3,000. As part of a series of calculations carried out by MPR to predict design effects due to clustering, it became apparent that halving the number of last-stage units (encounters) selected would make little difference to the overall effective sample size for the medical encounter abstraction component of the study, given the highly clustered nature of the design, and it would conserve data-collection resources.

MPR's goal in determining the number of users or encounters to select at each grantee was to arrive at approximately equal probabilities of selection. Had all grantee and site size estimates been equal (or close) to what was actually found to be the sample frame sizes, then equal numbers of users or encounters could have been selected for each grantee while meeting the goal of equal selection probabilities. Because

this was far from the case, the number to be selected in each grantee or site was adjusted up or down as appropriate (within logistically reasonable ranges) in an attempt to equalize the selection probabilities. An overall constant adjustment was made to the sample sizes for all grantees, in an attempt to maintain the overall targeted number of completes.

2. Grantee Sample

a. Definitions and types

A "grantee" is the administrative entity by which a community health center receives Section 330 funding from BPHC to provide primary medical care in the United States or its territories. Grantees can have one or more sites providing primary medical care services.

b. Eligible and ineligible grantees

All grantees in the contiguous United States that were operating and receiving Section 330 funding from the BPHC in 1994, and that had been operating and receiving such funding since calendar year 1992, were eligible for this study. The following CHCs were considered to be ineligible: those that are located in Alaska, Hawaii, or U.S. territories; those that were no longer operating or receiving 330 funding from the BPHC in 1994; those that had begun operating or receiving 330 funding after 1992 (or late in 1 992).[3] Ail sampled grantees met eligibility requirements.

Note that those grantees in operation during 1994 that began operations after 1992 (or late in that year) are not part of the target population for this study. Only those grantees that were well established by 1994 were considered eligible.

[3]The grantee had to have been in operation long enough to have submitted a Bureau Common Reporting Requirement report in January of 1993.

c. Coverage and multiplicity

Because the sampling frame excluded only those grantees that did not meet the eligibility criteria outlined in the previous paragraph there is no known source of noncoverage on the grantee frame. To the best of our knowledge, there were no grantees meeting the eligibility criteria that were not on the frame. Because it is believed that each grantee is represented on the frame only once, there is no known multiplicity (multiple chances for selection) among grantees.

d. Listing and selection

Fifty grantees were to be selected by the NCHS statistician out of a total of 501 eligible grantees on the frame, which was a January 1993 Bureau Common Reporting Requirement (BCRR) list containing all CHCs in the contiguous United States funded by the BPHC, minus those CHCs no longer receiving this funding in 1994. Selection was carried out with a probability proportional to size (PPS) methodology within stratum, based on the number of 1993 medical users reported by the CHCs on the January 1994 BCRR report, minus the reported number of migrant workers.[4] Nine strata were formed, based on Census region and urban/rural designation, along with a tenth for grantees with high proportions of managed-care patients..

The 50 selections were allocated across strata based on the proportion of the estimated number of users falling into each stratum (rounded to the nearest integer). The number of grantees to be selected per stratum ranged from two to nine. One grantee had a reported number of users larger than the sampling interval in its stratum, so it was designated as a certainty selection and became its own stratum, making a final total of 11 strata.[5]

---

[4]Migrant workers were defined as ineligible for the survey.

[5]It turned out later that this size measure was more than twice the grantee's actual size for this study, owing to its largest site's not being funded by Section 330 and thus being out of scope for this study.

To allow for substitution of an alternative grantee in the case of hard-core refusal at this first stage, the NCHS statistician formed grantee "clusters" within strata. Grantees were sorted in descending order of size within stratum, with consecutive pairs then forming the clusters, with two types of exceptions:

1. If such a pairing resulted in a certainty selection (that is, the combined cluster size was larger than the sampling interval), the larger grantee formed a cluster by itself.

2. If, at the end of the stratum, there were three grantees remaining, these three smallest grantees formed a cluster.

In this methodology, 250 such clusters were formed, from which 50 were selected via PPS systematic sampling. The NCHS statistician then selected one grantee PPS from within each selected cluster (through use of a random univariate between 0 and the cluster size), and the non-selected grantee within the cluster was the potential substitute. Two of the 50 selected grantees refused to participate, despite MPR and HRSA's efforts to gain their cooperation. The two substitutes also refused to participate, leaving 48 participating grantees in the study.

The grantee selection process had to be carried out twice because the first sample frame was found to have several problems: one grantee was ineligible, four eligible grantees were missing, and migrant workers had not been removed from the user counts prior to selection.

The probability of selection for each grantee (except for the one certainty selection, which has a probability of one) can be quantified as:

$$(1)\quad P(grantee) = \frac{x_h \hat{CMOS}_i}{\sum_{g=1}^{y_h} \hat{CMOS}_g} \cdot \frac{\hat{GMOS}_j}{\hat{CMOS}_i} = \frac{x_h \hat{GMOS}_j}{\sum_{g=1}^{y_h} \hat{CMOS}_g}$$

where:

x_h is the number of clusters selected in stratum h

y_h is the number of eligible clusters in stratum h

$\hat{CMOS}_i$ is the BCRR-reported measure of size (1993 medical users) for cluster I in stratum *h*

$\hat{GMOS}_{hij}$ is the BCRR-reported measure of size (1993 medical users) for grantee *j* in cluster I in stratum *h*

3. Site Sample

a. Definitions and types

A "site" can be thought of as a single location or facility at which a grantee provides health care services. A grantee may have only one site or may have multiple sites. Some multiple-site grantees have a main site with smaller satellite locations. Others have multiple sites functioning independently, with only a higher-level administrative tie to one another, in which case there is no main site. Some grantees have a combination of the two situations. Many grantees have school-based clinic sites, OB-GYN or other special-services clinic sites, or sites serving only certain types of patients (for example, women and children or the homeless).

b. Eligible and ineligible sites

All permanent sites (within selected grantees) that received Section 3 3 0 funding *and* provided primary medical services were considered eligible. (Note that, for some grantees, it was not easy to characterize whether a particular *site* received Section 330 funding, since such funding is given at the grantee level, with funds distributed in various ways across multiple sites,) Some sites were reported to be receiving Section 330 funding but not providing primary medical services, and some were reported to be providing such services but not receiving Section 330 funding. These situations had not been anticipated by HRSA or MPR. Such sites were considered to be ineligible.

Sites not receiving Section 330 funding or providing only specialized services (for example, substance abuse treatment, family planning services, or OB-GYN services) were considered to be ineligible, as were temporary "clinics" or "health fairs" for general or specialized health services (such as those set up at schools, senior centers, group homes, homeless or other shelters, mobile vans, hospitals, patients' homes,

or extended care facilities). In addition, sites providing services *only* to the homeless or only to migrant farmworkers were considered ineligible. Other sites providing services only to certain types of patients (such as women and children) were considered eligible as long as they provided general medical services.

School-based clinics were considered eligible if they provided comprehensive medical services. Such services were defined as providing 24-hour/year-round back-up (and allowing patients to use non-school-based clinics within the grantee system during vacations, holidays, and weekends); providing a full range of primary and preventive services; having access to laboratory and x-ray facilities; having procedures for hospitalizing patients, referring patients to specialists, and following patients through the system.

c. Coverage and multiplicity

Because we received the list of sites directly from each grantee, we believe that all eligible sites were covered in each grantee's site sampling frame. The only exceptions to this are two sites (in two different grantees) which opened late in 1994. The grantees were unable to provide information (such as the number of medical users) for these sites and would have been unable to produce a user sampling frame for these sites, had they been selected. For these reasons, the two sites were excluded from the site-subsampling process. No known multiplicity existed within these lists of eligible sites; each site is believed to have been listed only once.

d. Listing and selection

For each selected grantee, a grantee-provided list of all eligible sites was used as the frame for site subsampling. Sites were subsampled using a PPS systematic sampling methodology based on a 1994 site-level estimate (obtained directly from the grantee) of the number of medical users minus the number of migrant workers. Sites were sorted alphabetically by name for the selection process, and one, two, or three sites were selected per grantee, depending on the number and characteristics of sites under the grantee.

At the time the final decision on site subsampling was made, a prior set of decision rules had been in effect (involving the determination of whether patients would be willing to travel to sites other than their usual site, and the distance between each site and the "main" site), and many sites had already been subsampled under those rules (and hiring decisions made based on these selected sites). Under the newer rules, three grantees which had previously had two or three sites selected were to have only one selected; therefore, one site was subsampled from among those already subsampled.

Some grantees had great difficulty disaggregating counts of medical users for some of their sites. In two cases, the grantee provided us with one count representing two or three sites, in which case the aggregated sites were treated as one site for the purposes of site subsampling (unbeknownst to us at the time in one case).

The conditional probability of selection (given that the grantee has been selected) for a subsampled site can be quantified as:

$$(2) \quad P(site \setminus \text{grantee}) = \frac{s_j \, \tilde{SMOS}_k}{\tilde{GMOS}_j}$$

where:

s_j is the number of sites subsampled in granteej

$\tilde{SMOS}_k$ is the grantee-reported measure of size for site k in granteej

$\tilde{GMOS}_j$ is the grantee-reported measure of size for granteej.

Note that this probability can be (and was) greater than one for large sites in grantees having more than one site subsampled. Had any site been selected more than once, the number of users selected from that site would have been doubled (although this did not occur).

4. User Sample

a. Definitions and types

The sampling and analysis unit for the user survey is the medical user. A "medical user" is a person who had at least one medical encounter at the CHC. A "medical encounter" is defined by the BPHC as an encounter in which the person was seen by a physician, mid-level practitioner (nurse practitioner, nurse-midwife, physician assistant), or nurse.[6,7] It is possible to have more than one encounter in a single visit. Dental and other health encounters are not considered medical encounters, nor are services such as laboratory tests, x-rays, and prescriptions, unless a physician, mid-level practitioner, or nurse was also seen.

b. Eligible and ineligible users

For a medical user to have been eligible for this study, he or she must have had at least one medical encounter at an eligible selected site within a selected grantee in calendar year 1994. In addition, if the user had moved to a new residence since the last 1994 visit, he or she must have been currently maintaining a "usual" or "fixed' place of residence in the service area of the grantee from which he or she was sampled.'

By definition, anyone who did not have a medical encounter at an eligible selected site in a selected grantee in 1994 was not eligible for the user survey. In addition, medical users who were deceased, institutionalized, homeless, or who were migrant farmworkers at the time we tried to contact them were

---

[6]"The BCRR Manual" (1991), Bureau of Primary Health Care, Health Resources and Services Administration/Public Health Service, U.S. Department of Health and Human Services.

[7]Note that, for the visit study, a different definition of "medical encounter" was used. For that component, nurse-only encounters were excluded.

[8]Each site's service area is defined such that eighty percent of its users come from within that area.

all ineligible.[9] Migrant farmworkers were defined according to Bureau reporting definitions.[10] The homeless were defined as persons who either have no address listed in the CHC's records (meaning there is no way to contact them), or are reported by a family member, friend, or someone else to be homeless at the time of the survey.

Those who moved out of the CHC's service area since their last 1994 medical encounter were also considered to be ineligible. Students who used eligible school-based clinics but who had been given parental permission for only limited services (those typically provided by a school nurse's office) were out of scope for this study.

c. Coverage and multiplicity

There is known undercoverage of users in certain grantees for the following reasons. First, if the grantee had a separate OB-GYN site (there were five such sites in our sample, and they were considered ineligible for this study), then their obstetrics and gynecology patients had no possibility of being selected unless they also had a medical visit at an eligible site, whereas OB-GYN-only patients at grantees without such a specialty site could have been selected. Second, there were two grantees that each had a site that opened late in 1994 and could not provide counts for those sites (see "Site Sample" above). Any patients going only to those excluded sites in 1994 had no possibility of being selected. Third, one grantee could not provide a medical-user frame for the first half of 1994, because it had changed computer systems on July 1. Because of this, any patients making a medical visit only during the first six months of 1994 had no possibility of being selected. Note that the patients not covered in the sample frames for the last two

---

[9] An institutionalized patient known to be receiving care from the CHC was considered eligible.

"Migrant farmworkers and seasonal farmworkers are defined by the BPHC as individuals whose principal employment is in agriculture on a seasonal (as opposed to year-round) basis. Migrant farmworkers travel to a work area and live temporarily in the area while working there. Seasonal farmworkers work in the area of their permanent address and do not move temporarily to a work area. ("The BCRR Manual," 1991, BPHC/HRSA/PHS, DHHS). Note that seasonal farmworkers were considered eligible.

reasons are not believed to be substantively different from those that were covered by the grantees' frames. No other medical users are known to be under- or unrepresented by the grantees' user frames.

With respect to multiplicity, medical users who made a visit to more than one eligible grantee in 1994 had multiple chances of being selected, as did those who made a visit to more than one eligible site within a grantee. It should be noted that we asked those grantees with multiple sites selected to unduplicate their lists so that each listed medical user was represented only on one. There was also the possibility of frame problems in which a medical user could have been listed more than once on a given frame. Specific details on known multiplicity and corresponding weighting adjustments are found in II.C.4, "User Survey Weighting Methodology and Design Effects."

d. Obtaining lists of users and selection

The frames used to select the samples were paper or machine-readable lists of eligible users generated by the selected grantees from their databases. See "User Survey/Sample Implementation/Process and Procedures" for details on this process.

The number of users to be selected from each grantee was determined by a formula that attempted to attain equal probabilities of selection across all users while maintaining the overall goal of selecting enough users to obtain 2,000 completed interviews. Different formulas were used, depending on whether the grantee had sites subsampled or not. (See Appendix 1 for formulas.) The initial formula assumed an eligibility rate among selected users of 95 percent, and a response rate among eligible selected users of 76 percent, and adjusted for the fact that one grantee had refused to participate at that point in the process. When more than one site was subsampled, approximately equal numbers of users were selected from each site.

The sample size formula adjusted for the fact that the 1993 size estimates used in selecting the clusters and grantees (from January 1994 BCRR reports) and the 1994 size estimates used in selecting the sites (from the grantees themselves in early 1995) generally did not coincide with the actual 1994 frame sizes

from which the users were selected. In some cases, the estimated number of medical users was drastically higher or lower than what was found on the frame. Unfortunately, this caused the optimal sample sizes (optimal with respect to the goal of equal probabilities of selection) to be much smaller (n=15) or much higher (n=83) than predicted and initially presented to the interviewers (n=40). A decision was made to limit the sample size so that it was between 45 and 65 per grantee; that is, if the optimal number to be selected was less than 45, then 45 were selected, and if the number was greater than 65, then 65 were selected. Because, at the time this decision was made, user samples had already been selected for some grantees, two grantees required subsampling of users and three grantees required extra sample to be drawn. As a matter of course, replicate samples the same size as the original samples were drawn concurrently, in preparation for potentially low sample yields later on. The extra sample was drawn from these replicates.

After the number to be selected was resolved, MPR selected a systematic sample of numbers from each user frame. The grantees were then told which records on their frame(s) were part of the sample. Their frames were variously unsorted, or sorted by managed care status, sex, and/or prenatal care, as the capabilities of their systems permitted. We had requested that grantees sort their frames, if possible, by managed care status, sex within managed care status, and prenatal status within sex, to help ensure the proportional representation of these population subgroups in the sample.

A combination of grantee refusals, higher-than-expected ineligibility rates (closer to 10 percent than the previously assumed 5 percent, mostly due to people's moving out of the CHC's service area), high non-locatables, and frame sizes that were for the most part smaller than estimated caused the initial sample yield to be lower than targeted." Many discussions and calculations were carried out to determine whether to add to the selected number of users, and in what way. The final decision was to add sample to the

---

"On average, the actual number of users was about 80 percent of the estimate NCHS used to select the grantee, and about 90 percent of the estimate given to us by the grantee.

grantees in such a way as to even out the probabilities of selection, having in hand more information than was available at the time the initial samples were selected. A new formula was derived to arrive at an "optimal" sample size, this time assuming an eligibility rate among selected users of 90 percent (rather than 95 percent). (This was because, by then, a second grantee had refused to participate, and we used actual frame sizes, when available, to try to maintain the overall target of 2,000). (See Appendix 1.) Comparing this new optimal number to the initial optimal number, most of the grantees had additional samples of 5, 10, or 15 users selected from the replicate samples previously drawn. [12] New sample size cutoffs were set at 50 and 70, rather than 45 and 65, per grantee.

The conditional probability of selection for users selected from a grantee with no site subsampling (given the grantee has been selected) can be quantified as:

$$(3)\quad P(user \setminus grantee) = \frac{u_j}{GMOS_j}$$

where:

u_j is the total number of users selected from granteej

$GMOS_j$ is the actual user frame size for grantee j.

The unconditional probability of selection for such a user in stratum h (aside from any adjustments for multiplicity) is:

$$(4)\quad P(user) = P(grantee)\,P(user \setminus grantee) = \frac{x_h\ \hat{GMOS}_j}{\sum_{g=1}^{y_h} \hat{CMOS}_g} \cdot \frac{u_j}{GMOS_j}$$

The conditional probability of selection for users selected from a subsampled site (given that the site has been selected) can be quantified as:

---

[12] Making the additional sample sizes multiples of 5 allowed for multiple "harvests" from the replicate sample, which was divided into quintiles and randomly sorted within each quintile.

encounters at those excluded sites in 1994 had no possibility of being selected. Two grantees could not provide an encounter-level frame for all of 1994, since they changed computer systems mid-year (one on May 1 and one on July 1).[14] Because of this, any medical encounters during the first four or six months of 1994, respectively, had no possibility of being selected. Note that the encounters not covered in the sample frames for the last two reasons are not believed to be substantively different from those that were covered by the grantees' frames. No other medical encounters are known to be under- or unrepresented by the grantees' encounter frames.

With respect to multiplicity, there was only the possibility of frame problems in which a medical encounter could have been listed more than once on a given frame. There was no known multiplicity in any of the medical encounter frames; that is, we believe that each encounter was listed only once.

d. Obtaining lists of encounters and selection

The frames used to select the samples were paper or machine-readable lists of eligible medical encounters generated by the selected grantees from their databases. See "Visit Survey/Sample Implementation/Process and Procedures" for more details on this process.

Because it was expected that between-site within-grantee differences would be more significant for the encounter portion of the study than for the user portion, it was decided that no site subsampling should take place for the medical encounter abstraction part of the study. Although we asked for one frame per grantee, many ended up generating multiple frames (by site, by month, by date within site, etc.).

The number of encounters to be selected from each grantee was determined by a formula that attempted to attain equal probabilities of selection across all encounters while maintaining the overall goal of selecting enough encounters to obtain 3,000 completed abstractions. (See Appendix 1 for formulas.) The initial formula assumed an eligibility rate among selected users of 99 percent, and a completion rate

[14]Note that one of these two grantees was able to provide a medical user frame for the full year,

among eligible selected encounters of 95 percent, and adjusted for the fact that two grantees had refused to participate.

The sample size formula adjusted for the fact that the 1993 estimates of medical *users* were used in selecting the clusters and grantees (from January 1994 BCRR reports). Because encounter-to-user ratios differ across grantees, and because the user estimates were not always accurate, the optimal sample sizes (optimal with respect to the goal of equal probabilities of selection) ranged from 28 to 117 per grantee. A decision was made to limit the sample size so that it was between 40 and 80 per grantee; that is, if the optimal number to be selected was less than 40, then 40 were selected, and if the number was greater than 80, then 80 were selected. As a matter of course, replicate samples the same size as the original samples were drawn concurrently, in order to compensate for potentially low sample yields later on.

After the number to be selected was resolved, MPR selected a systematic sample of numbers from each encounter frame. The encounters were selected randomly within each sampling interval, so that any problems associated with periodicity that may have been present in the frame could be avoided. The grantees were then told which records on their frame were part of the sample. Their frames were variously unsorted, sorted chronologically, or sorted by site, as the capabilities of their systems permitted. We requested that grantees sort their frames chronologically, if possible, to help ensure the representation of the entire calendar year in the sample.

A combination of higher-than-expected ineligible rates (closer to five percent than the previously assumed one percent, mostly due to the unintentional inclusion of non-medical encounters on the frames) and frame sizes that were smaller than estimated, caused the initial sample yield to be lower than targeted." As with the user survey, a decision was made to add sample to the grantees in such a way as to even out the probabilities of selection, having in hand more information than was available at the time

[15]On average, the actual number of encounters was about 69 percent of the estimate given to us by the grantees. The average number of visits per user was about 2.8.

the initial samples were selected. A new formula was derived to arrive at an "optimal" sample size, this time using known frame sizes (when available) in an attempt to attain the goal of 3,000 completes. (See Appendix 1.) Comparing this new optimal number to the initial optimal number, most of the grantees had additional samples of between nine and sixteen encounters selected from the replicate samples previously drawn. New sample size cutoffs were set at 50 and 90, rather than 40 and 80, per grantee.

The conditional probability of selection for encounters (given that the grantee has been selected) can be quantified as:

$$(7)\quad P(encounter \backslash grantee) = \frac{v_j}{GNOV_j}$$

where:

v_j is the total number of encounters selected from granteej

$GNOV_j$ is the actual encounter frame size for granteej.

The unconditional probability of selection for an encounter in stratum h is:

$$(8)\quad P(encounter) = P(grantee)\,P(encounter \backslash grantee) = \frac{x_h\, \hat{GMOS}_j}{\sum_{g=1}^{y_h} \hat{CMOS}_g} \quad \frac{v_j}{GNOV_j}$$

e. Confidentiality

Consent and confidentiality were not issues of major concern for this component of the study because we did not ask the grantees to provide us with names and names were not recorded on the encounter abstraction forms.[16] We expected most grantees would want to maintain control over access to their records and do their own encounter abstraction, despite MPR's assurances of confidentiality. It turned out that roughly half preferred that our interviewer do the abstraction, owing to their staff's time constraints.

[16]Some grantees did provide us with names along with other sample information, even though this information was not requested. In those cases, the information was blacked out and not entered into any of our databases.

C. QUESTIONNAIRE DESIGN

1. CAP1 Questionnaire

a. Description of Content

The user survey questionnaire was administered as a computer-assisted personal interview (CAPI). Handcards with answer categories were given to respondents to read for 31 of the questions which appeared on the CAP1 screen. Approximately eighty percent of the questionnaire consists of questions taken from the National Health Interview Survey (NHIS) core or supplemental questionnaires. This is to facilitate comparisons with NHIS data. The remainder of the questions were developed for this survey and/or taken from other relevant surveys with validated measures for the variables not available from the NHIS. It consisted of eighteen modules.

Household Composition. This section establishes the family structure in the respondent's immediate household. Information was collected on each person in the household, including name, age, sex, and relationship to the sample person.

Language Usage. This section provides information on the possible effects of language barriers on the respondents' ability to get health care when sick, and to get appropriate preventive services (immunizations, mammograms, etc.). For preschool children, the questions were asked of the parent or guardian, who is likely to handle communication with providers. For sampled children who are school-age, however, the questions were asked about the child's language skills, because children often serve as interpreters for non-English-speaking parents.

Demographic and Economic Variables. This module contains questions about the demographic characteristics of the sampled person, including education, race, national origin, occupation (if employed), military service, and family income. These data will enable researchers to compare the health status and use of health services among the different demographic groups in the country.

Limitation of Activity, Health Indicator. This section provides information on the degree to which the CHC users are unable to carry out usual activities (work, school, housework) or are limited in the amount they can do, and on the health conditions that cause the limitation. It provides one indicator of whether the CHC population is more disabled than the general population.

Condition List. This section includes a checklist of chronic health problems, including heart disease, respiratory disease, digestive troubles, blood circulation difficulties, anemia, diabetes, problems with vision or hearing, etc. This section provides estimates of the prevalence of these conditions in the CHC population. The CHC population prevalence can be compared to that for the general population, to determine whether the CHC population has different or more serious health problems.

Chronic Disease Followup. This section collects information on whether common but serious medical conditions (hypertension, diabetes, asthma) are being appropriately monitored, and whether the CHC population is complying with physicians' orders. The survey also asked if tests and treatment were prescribed at the CHC. Most conditions in this section were framed within the referent period "ever."

Condition. This section provides information needed to assign a standard medical code to each condition reported in the Limitation of Activities section or the Condition List section. There are questions on the formal medical name for the condition, how long the condition has been present, what caused it, the part of the body affected, and other items needed for medical coding.

Disability. Adults were asked questions about their ability to perform normal tasks and the amount of assistance they needed with those tasks. These results will be compared to data for other populations. For children, questions were asked about any problems or delays in development.

Access to Care. These questions ask about the kinds of problems respondents may experience in trying to get access to health care. Questions were also included on the severity of the problem for which they were seeking treatment, and on whether health care was finally obtained. Respondents were asked reasons they were unable to get care when needed. CHCs were established to help eliminate access

problems for certain populations, so these questions will help to show how well this has been accomplished and identify remaining barriers to care.

Source of Care. This section contained questions about the places people go for care, their experiences with the care they received from tbe CHC, and their satisfaction with the care they received there. This data will help researchers to understand how people make decisions about where to go for care, what barriers to care they experience (for example, transportation, child care, language problems, cultural problems), and the areas where CHCs are doing well or could do more to reduce the barriers.

Routine Care. This section includes questions on routine checkups or physical exams, including time since last exam, what tests were performed, and whether counseling was given. The questions varied by age group, to determine whether age-appropriate tests and counseling were provided. For example, respondents were asked if, at their last routine examination, the physician or other provider inquired about diet, exercise, sexually transmitted diseases, drug and alcohol usage, emotional problems, etc. Finally, questions are included on other kinds of help or advice received from the CHC, to measure the extent to which the CHC system is helping users with a wide variety of problems.

Injury Control. This section asks about the kinds of precautions the CHC users take to prevent injury, including use of seat belts, smoke detectors, etc. For children, there were questions on the amount and kind of adult supervision to measure the risk of injury for these children.

Smoking and Pregnancy. For the general adult population, questions were asked about smoking history and current practice, and about the kinds of advice or help received from the CHC. This information will be used to measure the level of risk from smoking, which can be compared to the risk in the general population.

Because smoking during pregnancy is a particular concern, questions were included on that topic for sample persons who were pregnant at any time during the past year. (Some of the questions related to pregnancy were taken from the National Maternal, Infant, and Child Health Survey.) This section also

asked questions about other risky behaviors during pregnancy, and about pregnancy history, including prenatal care and adverse pregnancy outcomes. This information will be used to assess the outreach and counseling programs for pregnant women.

Cancer Screening. This section measures the extent to which CHC adult users are receiving appropriate cancer screening exams. For those users who tested positive to a screening exam, questions were included on problems encountered in getting appropriate follow-up tests and treatments. When age and sex appropriate, questions were asked about pap smears, rnammograms, proctoscopic, and other types of exams. Most of the questions were taken from NHIS, so that comparisons can be made between the CHC population and the general population.

Immunization. These questions asked whether recommended immunizations among pre-school children who use the CHC system are up-to-date. Immunizations include DTP/DT (Diphtheria, Pertussis, Tetanus), polio, MMR or measles, HIB or Hemophilus influenza, and Hepatitis B. Parents were asked to answer questions using their personal record of their child's immunizations. When this was not available, most questions were skipped, and the information was later abstracted from the child's medical record.

Health Behaviors. Questions were asked about a variety of behaviors (alcohol and drug use, exercise) and problems (depression, lack of food and money to buy food, etc.) that are related to health outcomes. This information will show whether the CHC population is at greater risk than the general population for adverse health outcomes, and where efforts should be directed. Respondents were also asked whether they had been tested for the AIDS virus (no questions were asked about test results).

Insurance. This section asked about health insurance coverage under both public (Medicare, Medicaid, other public assistance) and private plans. The information collected in other sections on health care and health outcomes will be compared for persons with and without health insurance, and with different kinds of coverage.

Income and Assets. Questions were asked earlier in the interview about total family income. This section identified the sources of that income; amounts were asked only for some forms of public assistance.

b. Process of Development

Prior to the issuance of the Request for Proposals (RFP), HRSA compiled questions for the user survey questionnaire, adapted primarily from the National Health Interview Survey (NHIS) core questionnaire and NHIS supplements over the last several years. Additional questions on health behaviors were adapted from the National Health and Nutrition Examination Survey (NHANES) III, the U.S. Department of Agriculture's 1994 Draft Supplement to the Current Population Survey (The Maternal and Child Health Survey), and the National Institute on Drug Abuse's 1993 National Household Survey of Drug Abuse. Original questions pertaining to community health centers and issues facing their users were formulated as well. A draft questionnaire was provided as an attachment to the RFP.

About two months after the contract was awarded, MPR received a revised version of this draft questionnaire from HRSA in machine-readable format, with each section having its own data file. MPR clerical staff then made an initial pass through the files, modifying each section of the questionnaire so that the questions and response choices were in standard CASES format." The re-formatted files were then passed on to MPR's programming staff

The first step taken by the programming staff was to familiarize themselves with the questionnaire layout. Any skip patterns, valid ranges/values, fills, and interviewer instructions that were already on the questionnaire were straightforwardly programmed into the CASES program. A "fill," a feature available in computer-assisted survey instrumentation, allows variable values to be inserted into the interviewer's script, depending on information already known about the respondent. For example, a question might start

"CASES is a computer program made available through the Computer Assisted Survey Methods Program (CSM), University of California, Berkeley. Neither the CSM staff nor the University of California bears any responsibility for the results or conclusions presented here.

with "Do you.. ." or "Does he.. ." or "Does she...", dependin$_8$ on whether it is a self-response interview or a proxy-response interview, and whether the proxy respondent is answering for a male or for a female. Both self and proxy versions of the questionnaire were developed simultaneously as part of a single instrument, making use of pronoun fills in this manner. Because the ordering of the sections was not determined until later on in the process, each section was initially dealt with independently.

One of the characteristics of computer-assisted survey instrumentation that distinguishes it from "paper and pencil" modes is that problems in the questionnaire, particularly those related to the skip logic, awkward question progression, and unnecessary repetition, become apparent more readily and much earlier in the process. Problems of this nature are not as obvious on hard-copy instruments, and so are occasionally not found until pre-testing or later, if at all. Much of the programmers' efforts at this point in the process focused on uncovering problems in and cleaning up the original instrument (questionnaire design tasks), rather than the CASES programming per se. All problems of this type that were discovered by the programmers were pointed out to the project staff who then worked with HRSA to resolve them.

As part of this effort, many skip patterns, valid ranges, fills, question wording, and interviewer instructions were corrected and new ones added to the questionnaire where necessary. Although not listed as such on the original instrument, it was decided that each question should have as potential responses both "don't know" and "refused." ("Don't know" was usually listed, but "refused" was generally not.) This necessitated that the programmers add skip pattern instructions for those responses which, even when listed on the original instrument, sometimes did not have skip patterns indicated.

Anotber time-consuming task for the programmers was determining the response-category codes used for comparable questions in the NIBS. This required looking through the hard copy NHIS core questionnaire and supplements as well as NHIS data on CD-ROM to find corresponding questions and their coding schemes. Having similar coding schemes is essential because the analytical plans for these data involve comparisons to national data collected through the NIBS. In addition, converting the hard-

copy instrument to a CASES instrument sometimes required breaking down a single question with multiple parts into multiple questions with branching logic.

In addition to making use of "hard range" checks (that is, not allowing any invalid or out-of-range responses to be entered), the computer-assisted instrument also allows for various types of consistency checks and "soft range" checks. For example, a consistency check would signal the interviewer to ask the respondent for clarification if earlier in the questionnaire (say, in the Condition List section), she stated that she never had hypertension, but is now saying (in the Chronic Disease section) that she has been told by a doctor that she had hypertension. In such a case, the program would prompt the interviewer to bring up the inconsistency to the respondent and repeat the question. Such checks are also used to identify situations where the interviewer mis-keys what the respondent just told her, in which case the interviewer will correct the response without having to reveal the inconsistency to the respondent. An example of a soft-range check would be if a person stated that he weighed over 500 pounds. In such a case, the interviewer would be prompted to read back the questionable response to the respondent and ask, "Is that correct?' These types of checks raise a flag to the interviewer, but allow seemingly inconsistent or invalid responses to remain if confirmed by the respondent. These "edits" were programmed into the CASES instrument to minimize the amount of editing that would have to take place after data collection ended.

Once the final order of the questionnaire was determined, the programmers were able to work on the more complicated part of the programming, namely establishing grids, creating macros, and compiling the conditions reported by each person. Grids are used to collect information on multiple family members, conditions, places, etc., when the number of such items will vary from respondent to respondent and a similar series of questions will be asked about each item. Macros allow the same computer algorithm to be repeated multiple times, generally over all items in a particular grid. Most complicated in this questionnaire was the compilation of a list of chronic conditions for each person, Because the questionnaire logic allowed a person to report a particular condition more than once, the programmers built

into the CASES program a check screen that asks the respondent whether the condition just reported is the same as any reported previously. In this way, the condition grid should contain each condition only once, while preserving the ability to map the condition back to any and all places it was originally reported. Similarly, the programmers built into the questionnaire a check screen to determine whether all household members had been enumerated by the respondent.

Once a preliminary version of the CASES instrument was ready, it was pre-tested by project staff, and it became apparent that the questionnaire was excessively long. The result was a decision to remove the section of the questionnaire that came from the NHANES on Eating Patterns, delete many of the questions from the Income section, and reduce significantly the section that asked about other family members' sources of health care. These sections were considered particularly burdensome and not essential analytically. In addition, for sensitivity reasons, new skip logic was added to the Smoking and Pregnancy section for women who had an adverse birth outcome. Minor modifications to the instrument itself (for example, refining the pronoun fills for the proxy version of the survey) continued until and through the interviewer training, where the practice sessions provided additional opportunities to pre-test the instrument. Two revisions to the CAP1 instrument were sent to the interviewers: one shortly after training, and another one about a month later. No modifications to the questionnaire were required after that point. (Only one minor skip pattern error affecting two questions on additional pap smears for about six respondents was discovered while the survey was in the field. The problem was resolved by contacting the affected respondents individually and modifying the program.) The final CASES program had over six thousand fields, covering about 3,400 questions.

Introduction/recruitment letters and the handcards used to assist respondents with questions containing lengthy or sensitive response categories were professionally translated into three languages: (1) Spanish (Mexican regionalism) (2) Cantonese/Mandarin and (3) Vietnamese. As sections of the user survey instrument were nearly finalized, they were professionally translated into Spanish (Mexican regionalism)

and then incorporated into a second CASES program. All subsequent changes to the English version of the survey were made in parallel to the Spanish version. Bilingual interviewers served as pre-testers for the Spanish version of the CASES program during training.

It was expected that all non-English-speaking respondents other than those who spoke Spanish would make use of an interpreter to carry out the interview. Using an interpreter to simultaneously translate this very long survey instrument proved to be particularly arduous for the Mandarin and Cantonese speakers. As a result, we had one of the bilingual interviewers prepare a hardcopy translation of the user survey into Mandarin.

C. Pretest

The pretest focused on both the user questionnaire and on the procedures and forms for the records abstraction. There were three distinct components of the pretest. First, the MPR project director and survey director attended a pretest of an early hard-copy version of the questionnaire at a CHC in Delaware (this was the version included with the request for proposal). This test was conducted by a HRSA contractor that had been responsible for developing sampling and operational procedures for this survey and for testing the questionnaire. The MPR staff observed the general complexity of the survey and recorded notes on respondent issues and problems. This effort occurred during the first month of the MPR contract and became an opportunity for staff to become familiar with the questionnaire.

The second phase of pretesting consisted of testing (at a nearby social services agency in New Jersey) a version of the hard-copy user questionnaire that was provided by HRSA shortly after the first pretest. The agency provided respondents of similar characteristics to the CHC population. The project director and another senior staff member conducted interviews with four female respondents. Respondents were retrospectively debriefed about any difficulties they had understanding or answering questions. This pretest substantiated the need to reduce the length of the questionnaire and also identified a number of

faulty skip patterns. It also provided another opportunity for senior staff to become more familiar with the detail of the user questionnaire.

The third phase of pretesting was the most substantial. Tests of the CAP1 version of the questionnaire were conducted at two CHCs, one in New Jersey and one in Delaware. The survey director and two field coordinators conducted pretests of the user questionnaire with seven respondents on two different occasions at the New Jersey site. Two men and five women were interviewed. The project director conducted cognitive pretests with two female respondents at the Delaware site. Results of these pretests were used to finalize the CAP1 questionnaire for interviewer training. The survey director also pretested the final CAP1 version of the questionnaire over the telephone with two respondents recruited from Delaware. Some general categories of pretest findings included:

- Additional faulty skip patterns were identified.
- Some programming errors were identified.
- Some respondents had difficulty answering some of the health behavior questions and alerted MPR staff to the need for privacy in conducting the interview. A screen asking parents of teenage respondents to leave the room before asking health behavior questions was added to the CAP1 program.
- Some questions were identified as upsetting to respondents. As a result, a decision was made to skip several questions for women with adverse pregnancy outcomes.
- Some questions required additional skip patterns.
- Some questions seemed inappropriately placed and were modified.
- Some respondents had difficulty reporting self-diagnosed conditions and impairments. Some additional probes were developed.
- The need for additional handcards was identified.
- Pretest interviews took over two hours, indicating the need to make additional cuts in the questionnaire. This was reinforced by respondent comments about the length and repetitive nature of the interview. As a result, the food frequency section was removed; the source of care section for other family members was reduced significantly; and the Income and Assets section was reduced.

- The single most problematic section was the one on physical limitations. Cognitive work on the NHIS had also indicated problems with this section. However, to keep it comparable to the NHIS, it was decided to leave the section unchanged.

- Fine tuning of some of the questions developed for this survey was also undertaken as a result of the pretesting.

The medical encounter abstraction form and procedures were also tested at the New Jersey site. The survey director and two field coordinators for this study abstracted several records using the draft version of the form, and staff at the CHC also abstracted several records. The form was also pretested by the previous HRSA contractor. Some of the problems indicated by the MPR pretest included:

- The absence of race and ethnicity in the record.

- The need to add additional categories for source of payment.

- The difficulty of reading physicians' handwriting.

- The necessity to consult with CHC staff to ascertain the type of provider in cases where the medical provider signed the record without indicating degree or position.

d. Recommendations

Pretests of the instrument and sampling and data collection procedures should be conducted in a larger number of CHCs.

After a preliminary examination of the data from the "Source of Care" section of the instrument, it was discovered that about one quarter of all respondents replied "no" to a question about whether they had been to any among a list of various types of health care facilities in the past 12 months. Another quarter said they had been to at least one of those places, but never mentioned the CHC. This meant that about half of the respondents missed a series of questions regarding why and how many times they visited the CHC in the past year, which were critical pieces of information. We recommend that, in the future, a verification screen be built into the instrument to validate cases in which the person claims to have not been to the CHC in the past year.

2. Medical Encounter Abstraction Form

a. Modeled after NHAMCS

One of the objectives of this study is to compare the data from the encounter survey with similar data for a nationally representative sample. Consequently, the instrument for this part of the study was modeled after the National Hospital Ambulatory Medical Survey (NHAMCS). The NHAMCS provides data for a broader-based population using a sample of ambulatory outpatient clinics within hospitals nationwide. While the visit survey was modeled after the NHAMCS, there are important differences.

The NHAMCS uses both retrospective and prospective data collection methods for patient visit data.[18] In the prospective method, the hospital staff collect the data at or shortly after the sampled visit has occurred. In this way, physicians and other relevant staff can be questioned if any of the requested information is unclear or missing. In the retrospective method, staff abstract data from medical records, but the time following the visit is likely to be considerably less than a year. In the CHC visit survey, all the data were collected retrospectively. Records were abstracted for the 1994 calendar year, and the abstractions were collected from July 1995 to March 1996. Compared to NHAMCS, retrieving missing information or clarifying an entry were more difficult and sometimes impossible.

As noted, all of the information collected in NHAMCS is done by hospital staff NHAMCS believes that hospital staff are better qualified to collect these data, as they are familiar with the record-keeping system, medical terms, and coding. In the visit survey, about fifty percent of the completed abstracts were done by grantee staff and about fifty percent by MPR abstractors. The type of collector used was determined by the needs and preferences of the grantee. Even when MPR staff did the collection, they consulted with grantee staff as needed. Finally, the abstraction form used in this study required the

[18] NHAMCS defines a visit as a direct personal exchange between an ambulatory patient and a physician or a staff member working under the physician's supervision for the purpose of seeking care and rendering personal health services. In the CHC visit survey, a medical encounter (the sampling unit) was defined as an encounter with a physician, nurse practitioner, nurse-midwife, or physician assistant.

completion of more data elements than did the NHAMCS form. The form will be discussed in detail below.

b. Operational Issues

Medical encounters were sampled from among all those occurring during the **1994** calendar year. A total of 3,000 completed abstractions was the goal for the study and 2,878 were completed. A medical encounter abstract form was to be completed for every medical encounter selected in the sample. The abstraction form was to be completed by either grantee staff or MPR staff, according to the grantee's preference. Grantees accepted the study's confidentiality procedures. Since there was only one MPR staff person at a given grantee, oversight by the grantee was made relatively easy.

Regardless of the collection method (grantee or MPR staff), MPR staff were available to answer questions, pick up and deliver forms, transmit forms to the home office, and assist with editing of the forms. When the local MPR staff member was not available or not able to answer questions, grantee staff were instructed to call MPR's home office.

c. Content and Instructions on the Form

The encounter abstraction form is in Appendix 2. As can be seen, information collected includes the following:

- demographics
- reimbursement type and source of payment
- patient referral type
- patient reported complaints and symptoms
- physician diagnosis, including any ICD-9 codes
- whether the patient was seen within the grantee system before
- medications and injections ordered or administered during the sample encounter

- disposition of the encounter
- providers seen at the encounter
- diagnostic and screening services received on the date of the sample encounter
- counseling, education, and other services associated with the sample encounter
- information on a second medical encounter if it occurred on the same day

Notable instructions included the following:

- the possibility that the reimbursement type and the source of payment would have to be determined from billing records and the possibility of having to use the "not available" box
- the importance of capturing the patients' complaints and symptoms in their own words, as distinguished from the physician diagnosis
- the need to note other chronic conditions in addition to the diagnosis for this encounter and the need to capture the ICD-9 code in the record
- the need to enter the name of all prescription and non-prescription drugs and the need to limit the number to ten by following decision rules
- the need to determine type of provider by the signature in the record when type is not specified
- the need to restrict the diagnostic and screening services provided to those provided on the same day of the sampled encounter

The abstraction form is different from the NHAMCS form in that it separates the reimbursement type from the source of payment; it adds the patient referral and past use boxes; it expands the diagnostic and screening entries; it expands the number of medications from five to ten; it modifies the entries for disposition to those more relevant for a CHC; and most significantly, it adds a long section on counseling, education, and other services. Other information remains identical on both forms.

d. Recommendations

Some of the information requested in the MPR abstraction form but not requested in the NHAMCS form was also more difficult to find in the record. This is particularly true of the counseling, education, and other services section. Moreover, because the information was abstracted up to 26 months after the actual encounter took place, certain information was not readily apparent in the record and could not be secured by the abstractors. One example of this is the listing of all types of providers seen at the encounter. Generally, because of these methodological differences, a full evaluation of the degree of missing or usable data by item should be assessed prior to another collection. Ideally, a small validity-and-reliability study should be undertaken to assess the quality of the information and the relative merit of including each item as designed in any future collection. Furthermore, an assessment should be made regarding the quality and completeness of data abstracted by the interviewer compared with data abstracted by CHC staff. Findings from such an assessment could be used as a basis of refining the training materials for any subsequent collections.

II. SURVEY IMPLEMENTATION

A. GRANTEE RECRUITMENT

1. Recruitment Process

Just after January 1, 1995, each of the executive directors of the 50 sampled grantees received a recruitment package containing an introductory letter and a copy of the study protocol (Exhibit II. 1). The letter, signed by the Director of HRSA, described the study and emphasized the importance of participation. It also set out the responsibilities of participating and the compensation that would be paid. The protocol summarized the details of the user and visit surveys, and addressed potential issues of concern from the perspective of the grantee. Each grantee was assigned one of three MPR field coordinators. The mailing was followed by a telephone call from the assigned field coordinator to the executive director to answer any questions or concerns.

The executive director was asked during that call to provide the following information:

- the number of clinics in the grantee system
- whether patients had a usual clinic, and if they did, how often they used a clinic other than their usual one
- whether the grantee's record system could identify certain patient demographic characteristics
- whether the records system could identify which patients were in managed care
- information on the nature of either the automated or manual recordkeeping systems
- contact person for site level information

Once this information was obtained, a questionnaire requesting information about each clinic site in the grantee system was either mailed or faxed to the designated contact person at each grantee. The

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

Health Resources and Services Administration
Bethesda MD 20814

EXHIBIT II.1

The Community Health Center (CHC) program has been providing critical health services to a significant segment of the American population for more than 25 years. To describe CHC users for both the Centers and the Bureau of Primary Health Care (BPHC) we have commissioned a study of CHC users. Of major interest are the health status and health needs of users; their health-related behaviors; and their patterns of health care utilization. These data will be used to assess the needs of the user population and to inform the planning process to ensure that the CHC program will meet users' future health care needs.

To complete this study successfully, we need your assistance. ***A sample of 50 CHCs across the country, including yours, has been selected to participate in this study.*** This study has two components. The first is a personal interview survey with a sample of users at your center, conducted by trained interviewing staff. The second component is a patient visit survey, in which data will be abstracted from medical records by contractor staff or by your staff if preferred for a sample of visits at your center. These two data collection components will be modeled after National surveys, so that the resulting statistics will be comparable to existing national data.

Your participation is critical to the study's success. This study is authorized by Section 241 of the Public Health Service Act (42 USAC 238j). Your participation is voluntary, and there are no penalties for not participating. However, the CHCs selected for the study have been carefully chosen to represent all CHCs in the country, and the cooperation of all selected CHCs is important for producing accurate, nationally representative data.

Enclosed is a general description of the study protocol and the role you may be asked to play. Your exact role in the study will be determined through discussions with our contractor, Mathematica Policy Research, Inc. (MPR). The health center and the survey participants will, of course, be reimbursed for the espenses of participation. In addition, participant confidentiality will be maintained, and each participant will be assured that there will be no adverse impact if they refuse to participate.

Our representative from MPR will be calling you in the next few days to discuss the details of the study and to determine the most efficient way to complete the study in your center. We recognize that this study will be an extra burden for you and your staff. However, with your help, MPR will develop procedures to minimize this burden. Since this is an important study, I strongly urge your full cooperation and thank you for your consideration.

Sincerely yours,

Marilyn H. Gaston
Assistant Surgeon General
Director

Enclosure

CHC USER AND VISIT STUDY

Overview of Protocol

The CHC User and Visit Study is a study of Community Health Center (CHC) users and medical care visits. It is being conducted in a national random sample of Community Health Centers. The study is conducted in two parts. One is the User Study which involves the collection of data concerning the health and demographic characteristics of CHC users. The second is the Visit Study in which patient visit data are collected describing patterns of health services utilization.

Data for the User Study will be collected through a personal interview with a sample of about 40 users in each CHC. The users sample will be randomly selected from among all CHC users who have had one or more visits to the CHC during the previous 12 months. This sample may be selected by computer in those CHCs having a suitable computer file of users. Those users selected in the sample will be sent a letter that briefly describes the study and asks for their cooperation. The letter will be followed by a telephone contact to verify the user's willingness to participate and to arrange as appointment for a personal interview. The personal interview will be conducted in the CHC or other convenient location by a trained interviewer provided by the HRSA contractor. The user questionnaire is a subset of the NHIS questionnaire, supplemented with questions relating directly to CHC users, and requires about 1.5 hours to administer. Users who participate in the study will be paid $20 to defray costs of transportation, child care and other expenses.

Data for the Visit Study will be obtained by abstracting information from medical records for a random sample of about 120 CHC visits. The visit sample will be randomly selected from among all visits made to the CHC in the previous 12 months. This sample may also be selected by computer in those CHCs having a suitable computer file of visits. For each visit selected in the sample, the patient's medical record will be pulled, and an abstract will be completed. The abstract is a one page form similar to the NHAMCS data form, and includes about 16 data items. Each abstract requires about 5-10 minutes to complete.

Mathematica Policy Research, Inc. under contract to HRSA, will have overall responsibility for data collection. Their staff will telephone and visit the CHC to arrange participation, collect data, and train and assist CHC staff in those study activities performed by the CHC. A number of study activities may be accomplished by either contractor staff or CHC staff, depending on the preference of the CHC. The following are the study activities that pilot tests have shown are generally accomplished most efficiently by CHC staff. These are also activities which many CHC administrators may prefer to have done by their staff rather that a contractor's staff. If the CHC prefers, of course, some or all of these activities may be completed by the contractor, with the advice and guidance of the CHC staff as needed. Final decisions on these and other aspects of the study will be made through discussions between the CHC administration and the contractor.

The study tasks which the CHC may consider performing are:

- Select a sample of users. With the contractor's assistance, select a random sample of "current" users according to specifications provided by the contractor. This is preferably done by computer when computerized tiles are available. Otherwise, the best available listing of users must be identified, and a sample selected manually.

- Mail a letter to each sampled user from the CHC administrator. A draft letter will be provided on diskette which the CHC may "personalize" as desired. The finalized letter must be addressed, printed on CHC letterhead, and mailed to each user in the sample.

- Telephone each sampled user and set up an appointment for the user to come to the CHC (or other selected location) for an interview. Recontact those users who do not appear at the appointed time to arrange a second appointment.

- Provide space in the CHC or help arrange for available space at a nearby location for conduct of personal interviews with the users. The space must be reasonably private to permit confidential interviewing.

- Select a sample of patient visits. With the contractor's assistance, select a random sample of visits according to specifications provided by the contractor. This is preferably done by computer when computerized files are available. Otherwise, the best available listings or logs of visits for the previous 12 months must be identified, and a sample selected manually.

- Pull the medical records for the sample of patient visits. Abstract about 16 data items from the medical record for each sampled visit. In CHCs with computerized information about each patient visit, some of the visit data may be derived from the computer files.

number of sites ranged from one to more than eighteen per grantee. The information asked for each site included the following:

- whether or not Section 330 funding was received
- whether or not general medical care was provided
- the number of medical visits for 1994
- the number of medical patients seen in 1994
- basic demographic characteristics of patient population
- percentage distribution of languages spoken by patients
- driving time from the site to the central site
- the perceived willingness of patients to travel to another site to be interviewed
- the willingness to have staff at the site abstract information from records for the visit survey

The information from these two information collections was critical to completing the sample design and developing the operational details for sampling and interviewing. Many grantees had considerable difficulty reporting this information and often had to contact each site to determine the answers. There were problems with the reliability and the accuracy of the reported information, as well as with its timeliness. This was the case despite the considerable effort and cooperation of the grantees. Some of these problems will be discussed in more detail later in the chapter to facilitate the planning of any future surveys.

2. Grantee Description and Participation

Fifty grantees were sampled. Forty-eight agreed to participate and two refused. For each grantee, the sample design provided one substitute which, in case of a refusal, could be recruited in its place, but each substitute for the two refusing grantees also refused. The reasons for refusals were related primarily

to limited staffresources during the planned data collection period. In one case it was related to a concern about the confidentiality of the data.

As Table II. 1 indicates, ten of the 48 participating grantees (2 1 percent) had one eligible clinical site (CHC) and 3 8 (79 percent) had multiple sites. A little over 60 percent of the grantees had between two and five eligible sites, and 19 percent had between six and twelve eligible sites. To be eligible, a site had to receive Section 330 funding and had to provide general medical care to the patients. Some sites did not meet the criteria, and some grantees had difficulty reporting which sites were eligible for the study. Sites often received multiple sources of funding, and in some cases the grantee had one main site, with the other sites functioning as smaller satellites of the main site. In others, grantee sites were grouped together for administrative and/or political reasons, and functioned, for the most part, independently.

Among the 48 participating grantees, ten had only one site and six had two sites with no subsampling of sites. Table II.2 indicates the number of eligible sites that were subsampled from grantees with two or more eligible sites. Twenty-four grantees had one site subsampled, three had two sites subsampled, and five had three sites subsampled.

As Table II.3 indicates, **31 (65** percent) of the 48 participating grantees sampled identified the executive director or the acting executive director as the primary contact for the entire study, including the grantee level information collection. Seven (15 percent) of the grantees identified an operations or clinic director. The remaining grantees identified a variety of contact points.

3. Timeframe for Recruitment

We initially planned for recruitment to take place over six weeks and, with very few exceptions, the agreement to participate was given within this period. As Table II.4 indicates, all the participating grantees agreed to participate within the planned six weeks of the recruitment process, except one grantee that waited almost four months before consenting. Of the refusals, the last grantee did not clearly refuse until almost five months after the initial mailing. Interviewers were trained centrally at the end of March. The

TABLE II. 1

NUMBER OF ELIGIBLE SITES PER GRANTEE

Number of Section 330 Sites at the Grantee	Number of Grantees[a]	Percent of Total	Cumulative Percent
1	10	20.8	20.8
2	12	25.0	45.8
3	8	16.7	62.5
4	5	10.4	72.9
5	4	8.3	81.3
6	2	4.2	85.4
7	1	2.1	87.5
8	3	6.1	93.7
9	1	2.1	95.8
10	1	2.1	97.9
11	0	0	97.9
12	1	2.1	100.0
Total 172	**48**	**100.0**	**100.0**

"Two original sample grantees and their sampled substitutes refused.

TABLE II.2

SITE SUBSAMPLING IN THE USER SURVEY

Number of Eligible Sites	Number of Sites Subsampled	Number of Grantees
1	1[a]	10
2	2"	6
2	1	6
3	2	2
3	1	6
4	2	1
4	1	4
5	3	1
5	1	3
7,8,9,12	3	4
6,8,10	1	5
172	67	48

"No subsampling took place.

TABLE II.3

PRIMARY CONTACT IDENTIFIED FOR ALL SAMPLED GRANTEES

Primary Contact	Numbers	Percent
Executive Director or Acting Executive Director	31	65
Operations or Clinic Director	7	15
Deputy Director	1	2
Executive or Administrative Assistant	3	6
Patient Services Director	1	2
MIS Director	1	2
Financial Director	1	2
Title Not Clear	3	6
Totals	**48**	**100**

TABLE II.4

DATES FOR GRANTEE RECRUITMENT

Ending Date of Week Recruited	Number of Grantees Recruited	Percent
A. Participants		
01/06/95	6	12.5
01/13/95	29	60.4
01/20/95	11	22.8
01/27/95	--	--
02/03/95	1	2.1
02/10/95	--	--
02/17/95		
02/24/95	--	
04/25/95	1	2.1
Total	48	**100.0**
B. Refusals[a]		
01/17/95 (Original)	1	--
05/16/95 (Substitute)	1	--
01/27/95 (Original)	1	--
02/14/95 (Substitute)	1	--

[a]Two grantees and their respective substitutes refused to participate.

late decisions of grantees to participate or withdraw from the study required us to make adjustments in staffing and site-specific sampling plans.

4. **Problems with Grantee Recruitment and Reporting**

For the most part, grantee recruitment went smoothly. Individual administrators, in consultation with their staffs, decided whether or not to participate in the study. In two instances, because of a change in executive staff, administrators required a decision from their Boards of Directors. The timing of the recruitment largely went as planned. One of two exceptions was a late refusal where an interviewer was already trained and waning to begin; another arose when the start of interviewing was delayed by over a month following training.

While recruitment itself was not a problem, the collection of information about site characteristics was unexpectedly difficult. The site level information requested on the grantee questionnaire was difficult for grantees to provide. Many items were left blank, and other information provided was either incorrect or outdated, particularly when MPR staff obtained information from more than one source at the grantee. Collecting and arranging for the sampling for the user survey was most problematic. The user sampling took almost six months to complete for all grantees, although the vast bulk was completed during April 1995. The encounter sampling took about eight months from start to finish, although most grantees' encounter samples were selected during June 1995.

While validating information does not exist for all measures, certain data originally reported by grantees was later confirmed through a more accurate process associated with developing the sampling frame. Specifically, estimates originally reported for the number of medical users of the sampled sites in 1994 and estimates for the number of medical visits in 1994 across all a grantee's sites can be compared with those later derived in the development of the sampling frame. This comparison gives some indication of the reporting difficulties for information collected at the grantee level. The information originally reported was used for the refinement and planning of the sample design, as discussed in detail in the

sampling section; large reporting errors necessitated modifying the sample design to account for the actual caseload size in the sites.

Table II.5 shows the difference between the number of 1994 medical users and visits reported initially compared with the number subsequently reported as the sample frames were developed. For users, only 47 percent of the sites provided an estimate within 10 percentage points of the frame size. For visits, the figure was 38 percent. These large discrepancies required us to make substantial adjustments to the sample selection plan.

Accurate information on the number of eligible sites and the distances between them was essential to develop the survey operations plan. The mean driving time between a site and the central administrative location is 27 minutes. As Table It.6 indicates, 3 8 percent are within 15 minutes; 22 percent are between 16 and 30 minutes; 26 percent are between 3 1 and 45 minutes; 5 percent are between 46 and 60 minutes; and 8 percent are over 60 minutes. These estimates of travel time between sites and the central office do not address the travel time between sites, which was over two hours in a few cases. Better information regarding the number of eligible sites and the travel time between sites will improve the planning and budgeting for future surveys, especially if data are to be collected in every site.

5. **Recommendations**

The problems we encountered as part of the grantee recruitment process are unfortunately unavoidable. Our recommendation is to recognize that the process is time consuming and labor intensive. Substantial resources and time should be planned for this phase in any future study.

TABLE II. 5

PERCENTAGE POINT DIFFERENCE FOR ESTIMATES
OF NUMBER OF USERS AND VISITS
(Number of Grantees)

	Users		Visits
Percent Difference Between	Grantee-Reported Users in 1993 (BCRR) and Grantee-Reported Users in 1994 (Grantee Interview)	Grantee-Reported Users in 1994 (Grantee Interview) and User Survey Frame Size[b]	Grantee-Reported Visits in 1994 (Grantee Interview) and Visit Survey Frame Size
-80 to -61%	0	0	2
-60 to -3 1%	4	9	7
-30 to -11%	11	18	20
-10 to -6%	4	1	-4
-5 to -2%	5	10	3
Within 1%	6	10	7
2 to 5%	6	4	2
6 to 10%	5	3	2
11 to 30%	2	3	1
31 to 60%	5	1	0
⋮			
111 to 120%	0	1	0
Overall Percent	-14%	-12%	-31%

N.B A negative percent means the newer size estimate was lower than the older one. The percent difference is calculated relative to the older size estimate.

[a]For grantees in which site subsampling took place, this estimate was at the site level, rather than the grantee level.

'There are a total of 60 frames that are reflected in this column -- 15 one-site grantees (one such grantee never gave us an estimate); 24 multi-site grantees with one site selected; and 21 sites in 8 grantees in which two or three sites were selected.

TABLE II.6

DRIVING TIMES FROM SITE TO THE CENTRAL SITE

Minutes	Number of Sites	Percent of Total
1-15	37	38.5
16-30	21	21.9
31-45	25	26.0
46-60	5	5.2
Over 60	8	8.3
	96	100

"Driving times were available for 96 eligible sites.

B. TRAINING

1. User Survey

A five-day centralized training session for the User and Visit Surveys was held at MPR's home office near Princeton, New Jersey, in March 1995. Fifty-one interviewers were trained: one interviewer for each originally sampled CHC that agreed to participate at the time, and two floating, backup interviewers. The training was attended by one of the HRSA project officers. An additional day would have been required had the encounter survey training been conducted in full. Instead, the encounter survey procedures were briefly covered and the form introduced, but subsequent training was conducted by telephone to accommodate design changes and delays in obtaining the encounter sample from CHCs.

The training covered the study methodology; detailed user survey procedures; general interviewing; general CAPI training; an overview and question-by-question specifications for the user questionnaire; and administrative information. A detailed agenda (Exhibit II.2) outlines the training, which consisted of large-group lectures and exercises, and small-group practice sessions with laptop computers. Community role plays were used to demonstrate and practice the user questionnaire in the large group, and both community and dyad role plays were used in small-group sessions. A staff-to-interviewer ratio of 1:6 was maintained in the small-group sessions.

The challenge of this training was the user questionnaire, which consisted of more than 3,400 questions and generally took between one and a quarter and two hours to administer. In addition, there were multiple versions of the questionnaire to be trained on, including self and proxy adult versions; a self and proxy teenager version; an adult-reporting-for-a-child version; and English and Spanish versions. A comprehensive training manual was developed, with chapters following the agenda topics. Small groups were arranged by expected CAP1 skill level, and trainers were assigned accordingly. Trainers were assigned in pairs, with one strong in CAP1 and one in questionnaire design. In addition, every small group

EXHIBIT II.2

NATIONAL STUDY OF COMMUNITY HEALTH CENTERS
USER AND VISIT SURVEY

AGENDA

MARCH 26 THROUGH MARCH 30, 1995

SUNDAY (March 26.1995)	**Approximate Times**
INTRODUCTIONS	
MODULE 1: • Overview and Background • Survey Methodology and Mode • Interviewer's Role	8:30-10 a.m.
MODULE 2: • Study Definitions and Terms • Overview to Sampled CHCs (Grantees) • Sampled CHC Sites • Eligible Respondents	10-11 a.m.
MODULE 3: • Overview to General Interviewing • Probing	11-12 noon
LUNCH	
MODULE 4: • Caring for and Using the Computer • Introduction to Function Keys • Solving Problems • Skip Sequences	1-3 p.m.
COMMUNITY ROLE PLAY (ADULT SELF RESPONDENT)	3-5:30 p.m.

EXHIBIT II.2 *(continued)*

MONDAY (March 27.1995)	**Approximate Times**
MODULE 5:	
• Overview of the Questionnaire	8:30-9:30 a.m.
• Review of Q x Q Main Points	9:30-12:00 noon
LUNCH	
SMALLER GROUP PRACTICE	1:00-5:30 pm
• Mock #1	
• Mock #2	

TUESDAY (March 28.1995)	**Approximate Times**
MODULE 6:	8:30-9:15 a.m.
• Confidentiality Procedures	
• Confidentiality Pledge	
• Consent Issues and Forms	
MODULE 7:	9:15-11:00 a.m.
• Contacting the Administrator	
• Contacting the Respondent	
• Common Questions and Answers	
MODULE 8:	11:00-12:00 noon
• Monitoring and Verification	
• Timesheets	
• Expense Forms	
LUNCH	
SMALLER GROUP PRACTICE	1:00-5:30 p.m.
• Mock #3 - Dyad	
• Mock #4 - Dyad	

WEDNESDAY (March 29.1995)	**Approximate Times**
MODULE 9:	8:30-9:45 a.m.
• Identification Numbers	
• Contact Forms	
• Status Codes	

EXHIBIT II. 2 (continued)

MODULE 10: • Weekly Production Report • Transmitting Questionnaire Data • Transmittal Forms • Mailing Instructions • Incentive Receipt Forms	9:45-10:45 a.m.
MEETING WITH FIELD COORDINATORS • Site Specific Information • Arranging Reporting Appointments	10:45-12:00 noon
LUNCH	
SMALLER GROUP PRACTICE • Mock #5 • Open Practice	1:00-5:30 p.m.
PRACTICE: Spanish Version	Evening
THURSDAY (March 30.1995)	**Atmroximate Times**
LARGE GROUP MEETING • Spanish Version • CAPI Version	8:30-10:30 **a.m.**
MODULE 11: • Introducing the Visits Abstraction Form • Procedures	10:30-12:00 noon
LUNCH	
LARGE GROUP MEETING AND PRACTICE	l :OO-3:00 p.m.

had a few relatively strong or experienced interviewers, and remedial sessions with more individualized attention were offered at night.

Follow-up monitoring and practice occurred once the interviewers returned home, before they began the actual interviewing.

2. **Visit Survey**

The training for the survey of medical encounters involved training both MPR interviewers and CHC staff, as some CHCs elected to have their own staff conduct the records abstractions. The encounter abstraction form was introduced to the interviewers at the centralized training, but no formal training was conducted. Telephone conference training sessions were conducted by MPR staff for separate groups of five to six MPR abstractors. This follow-up training was scheduled after the user survey was well underway at each CHC. A separate training manual was prepared for the encounter abstraction. This manual presented a review both of transmittal and editing procedures, and included question-by-question instructions and definitions for completing the abstraction form. The manual provided the basis for assuring consistency in the collection of information.

Telephone training for CHC staff members was also conducted in groups of five or six, whenever possible. Individual training was arranged for CHC staff who were not available for group training. For CHC staff, an abbreviated version of the manual was prepared which excluded the transmittal procedures, since transmission of forms was to be handled by the assigned interviewer. CHC-specific information was appended to each manual. Staff at MPR were available to answer questions as both CHC staff and interviewers began the abstraction process.

Exhibit II.3 displays the agenda used for the records abstraction training.

EXHIBIT Il.3

VISIT SURVEY TRAINING

Two Hour Session

Introduce Participants in Training

General Background Information

Overview of Materials Provided for Completing Visits Survey

- Training Manual
- Abstract Forms
- Contact Sheets
- Log (submit copy in lieu of Visits Survey batch sheet)

Question-by-Question Review of the Abstract Form

Review of Procedures

- Gaming Cooperation from Medical Records Personnel
- Accessing Medical Records
- Obtaining Information from Billing Records
- Editing Completed Forms
- Submitting Forms to MPR

Scheduling for Completing Remaining Work

C. USER SURVEY

1. Sample Implementation

a. Process and procedures

All grantees were sent instructions (see Appendix 3) on how to create the sample frame of medical users. The first step was to read the instructions carefully and then call the assigned field representative at MPR. Single-site grantees were instructed to compile a list of all persons who made at least one medical visit to their eligible site in calendar year 1994. Similarly, for two-site grantees in which no site subsampling took place, the grantee was instructed to compile one list of all persons who made at least one medical visit to either of their eligible sites in 1994. If one or more sites were subsampled, the grantee was instructed to compile such a list separately for each selected site. To the extent possible, all grantees were asked to remove ineligible users, such as those who were known to be migrant farmworkers, homeless, or deceased, or known to have moved out of the service area, and to remove any duplicate entries of medical users. The lists of eligible medical users were generally in machine-readable format, although some grantees' systems could generate usable lists only on paper. A handful of grantees compiled their user lists completely manually. Once the list(s) were compiled, the grantee was instructed to call the MPR field coordinator once again to give the total number of records on the list(s).

At this point, the field coordinator gave the frame size(s) to the Sampling Manager, who then systematically selected the user list numbers for the sample. These numbers were then sent by fax to the grantee. Previously, either at the point the list was first generated or after the frame size was given to MPR the grantee was asked to sort (re-order) the list by managed care status, sex (within managed care status), and prenatal status (within sex), to the extent its system had this capability. The grantees were also instructed to number their user lists consecutively, beginning with the number one. When the grantees received the selected user list numbers, they were asked to provide as much of the following information as was available on each of the users corresponding to the selected list numbers: list number, name,

address, telephone number, date of birth, sex, medical record number, principal service site used, parent's name (if under age 18), and date of last CHC visit.

As described more fully in the Sample Design section, user sampling took place in two stages. The initial sample was supplemented by additional sample later in the field period, owing to yields that were lower than expected. When the additional sample was selected by the Sampling Manager, the selected numbers were sent via fax to the grantee, who was then instructed to go back to the same list(s) of eligible medical users and provide the contact and other information on each of the users corresponding to the selected list numbers.

Once the contact and other information was received for the sample, the cases were entered into a database, assigned study identification numbers, and set up on the receipt and tracking system; then $20 respondent incentive checks and interviewer log sheets were generated.

In this manner, a total of 2,964 cases were selected: 2,649 in the initial sample and 3 15 in the additional sample. At the end of the field period, grantees were paid $250 for their time in preparing the sample lists and participating in the study. Eleven of the grantees did not create the user sample frames themselves, but used the services of their data processing vendor. Nine of these eleven grantees had a single vendor, who did not charge for the service. The remaining two vendors charged fees of $200 and $300 for their services. When a vendor provided the sample frame for a fee with only negligible effort on the part of the grantee, the grantee did not receive the $250 fee.

b. Timeframe

User samples were selected on a flow basis immediately after each frame size was received from the grantee. The first user sample was selected on March 29, 1995, just before the interviewers were trained. All but four of the initial user samples were selected during April and early May of 1995. Two grantees had their initial samples selected in late May, another in early June, and another in late September of 1995.

The additional user samples were selected on August 1, 1995, for all sites, with the size of the additional sample for the last site being estimated.

For many of the grantees, obtaining the user frame counts took much longer than anticipated. Some had staff shortages, others had systems limitations (for example, they had to run the program to generate the list overnight or on weekends so as not to interrupt daily systems tasks), and still others began to have second thoughts about their participation in the study. Staff turnover at all levels was a major problem for a surprising number of grantees. When a delay occurred in the sampling process, it was often due to the departure of a staff member who had been responsible for assisting, either directly or indirectly, with sample selection. Some grantees had to generate the frame more than once because the first one was generated incorrectly. For some of the nine grantees served by the one data processing vendor, even obtaining the authorization for the vendor to proceed took an extended period of time. Once the frames were generated and the samples selected, there was often a long delay for some grantees to provide the corresponding contact information for the selected cases.

Five of the grantees who had additional sample selected could not provide contact information for their selected cases in a timely manner. Because we wanted to close down the field period, it was decided in mid-October 1995 that these 3 5 additional cases should not be released and hence should not be considered part of the selected sample. That left 2,929 medical users selected in the sample.

c. Problems

In addition to the problems mentioned above and in the Sample Design section, there were others worth noting. Many grantees could not sort their frames by any of the three specified data items (managed care status, sex, or prenatal status) and some could not consecutively number their lists. When unable to number, they found a way to find the selected cases, such as counting the number of cases per page or screen, locating the page or screen on which the selected case could be found, and then counting out which record on the page or screen was the selected one. None could remove all types of ineligibles from their

frames, but most could remove some types of known ineligibles (or did not have such patients on their systems to start with); however, three grantees had particular difficulties compiling the type of list required. The first was unable to remove dental-only users from the frame. The second was unable to limit the frame to 1994 users only. Because we knew about these problems before the samples were selected and had some idea of the proportion ineligible, a proportionately larger sample size was selected. A total of 149 ineligibles were found from these two grantees because of these frame problems. A third grantee (which had separate frames for the first seven months of 1994 versus the last five months owing to a mid-year systems change) could not separate out users from non-selected sites from their January-through-July frame. A slight oversample (yielding about six extra cases) was selected for this part of the frame, although it turned out that no users from non-selected sites appeared in the sample.

Other types of problems stemmed from grantees' antiquated and/or inflexible computer systems. Several grantees upgraded their computer systems mid-year, usually completely changing systems (and data processing vendors, when used). When this happened, most grantees tied to piece together a frame for the entire year, but one grantee could provide a frame of medical users only from July through December of 1994, leaving off any medical users who made a visit only in the first half of the year. Some were put in the awkward position of dealing with data processing vendors that no longer provided service to them, but who had provided service in 1994. One grantee did not have internal computing staff who could construct the sample, and was initially reluctant to let us deal with its data processing vendor. At some grantees, there were selected medical users who stated that they did not ever visit the community health center, or at least did not make a visit in 1994. The contact information (address, telephone number) for many grantees was incomplete or out of date.

Of the 2,929 selected medical users, 149 were ineligible because of known frame limitations (as mentioned above). Another 217 cases were found to be ineligible after an attempt was made to contact them, and 392 cases were nonlocatable and hence of unknown eligibility status.

Three users from three different grantees were actually selected twice in our sample. The duplicate selection is considered an ineligible case. Two of these three grantees insisted that their sample frames were unduplicated except for the one case that was actually selected twice: one was able to explain why the person was selected twice (he had a special account) and confirmed that no other selected cases were in that situation; the other was unable to explain why the person was on the frame twice. The third grantee that had a user selected twice was able to explain the cause of the multiplicity: all users of their school-based clinics were also on the patient registration files for their main site. It was discovered that a fourth grantee had users listed more than once on its frames, but no users were actually selected twice. This problem occurred when a patient was inadvertently re-registered into the system under a different identification number. No other grantees admitted to having a problem with duplicate listings.

d. Recommendations

For any similar endeavors carried out in the future, these problems are less likely to be encountered because grantees are upgrading their computer systems, thereby enhancing their capabilities to generate sample frames. Because there was such little control over the frame development process, one recommendation would be for the grantees to supply the contractor with the sample frames, and have the frame examined, unduplicated, and sorted by the contractor prior to sample selection. In this way, both the quality and the timing of the sample selection process would be improved.

2. CAPI Questionnaire

a. Fielding the CAP1 Questionnaire

Confidentiality and consent. Sampled users were sent a letter from the grantee, letting them know that someone from Mathematica Policy Research would be contacting them regarding this survey and

encouraging them to participate.' Exhibits II.4.a through II.4.c are prototype advance letters sent to grantees. Exhibit II.4.a is for adults and teenagers who were sampled. Exhibit II.4.b is for the parent of a teenager who was sampled. Exhibit II.4.c is for the parent of a child who was sampled. The users were assured that their responses and identities would be held in strict confidence. During training, the interviewers were informed about the various confidentiality laws in effect and the civil and criminal penalties that would result should they violate the confidentiality of the survey respondents. They all signed agreements confirming their understanding of the confidentiality requirements and the penalties for any violation. While all contacted users were strongly encouraged to participate, all users had the right to refuse.

There was the possibility of a future validation study of the responses given by the respondents in the user survey. Such a study would have required a look at each patient's medical chart, so we asked all respondents at the end of the interview for permission to obtain their medical record. Ninety-two percent of respondents signed the consent form. (Such a validation study did not take place.) In addition, for sampled children under age six, if the parent or guardian responding to the interview on behalf of the child did not have a personal record of the child's immunizations available during the interview, he or she was asked to sign a consent form granting permission to look at the child's medical chart to determine immunization status. Among the 122 children in this situation, 95 percent of the parents/guardians gave written permission to do so.

Contacting Respondents. Interviewers contacted sample members by telephone, when feasible, to arrange for an in-person interview. When a sample member could not be reached by telephone an in-person visit was made to either conduct the interview or arrange for a more convenient time.

'For one grantee, it was not feasible to send letters prior to contact because the grantee did not provide sample information until very late in the data collection period.

EXHIBIT II.4.a

GRANTEE LETTERHEAD

INTRODUCTORY LETTER TO CHC USER AGED 13 AND OLDER
(ADULTS AND TEENAGERS)

Dear _____

I am writing to ask you to take part in a study of people who use Community Health Centers. The study is being conducted by the U.S. Public Health Service, which supports Community Health Centers. Our center, like others across the country, provides health care to people with many different health care needs. Learning about those needs, and the ways that people use health centers is very important for planning the future of the health center program.

This study is being conducted in many health centers across the country. You have been randomly selected to represent the users of our center in this study. We are asking you to complete an interview at the community health center with a specially trained interviewer. Questions will be asked about your health and health care needs. The interviewer will also ask about all of the places you go for health care, and opinions on the care received. For completing the interview you will receive $20 to cover any expenses.

Participation in this study is voluntary. If you decide not to participate, it will have no effect on the services received at our center. However, your cooperation is very important for making the study a success and for helping the center understand the health care needs of its users. All information will be used only for statistical purposes. All answers will be held confidential. The interviews usually take about an hour and a half.

In a few days, a staff member from our center will call you to answer any questions you may have, and to arrange for an interviewer to call you. In the meantime, if you have any questions about the study, please call us at the following telephone number: _____

Sincerely,

This study is authorized by Congress in Section 241 of the US Public Health Service Act (42 USC 238j).

EXHIBIT II.4.b

GRANTEE LETTERHEAD

INTRODUCTORY LETTER FOR PARENT OF CHILD AGED 13-17 YEARS

Dear _____

I am writing to let you know that _____ has been selected to take part in a study of people who use Community Health Centers. The study is being conducted by the U.S. Public Health Service, which supports Community Health Centers. Our center, like others across the country, provides health care to people with many different health care needs. Learning about those needs, and the ways that people use health centers is very important for planning the future of the health center program.

This study is being conducted in many health centers across the country. Your child has been randomly selected to represent the users of our center in this study. We will be asking your child to complete an interview at the community health center with a specially trained interviewer. Questions will be asked about your child's health and health care needs. The interviewer will also ask about all of the places your child goes for health care, and opinions on the care received. For completing the interview your child will receive $20 to cover any expenses.

Participation in this study is voluntary. If your child decides not to participate, it will have no effect on the services received at our center. However, your cooperation is very important for making the study a success and for helping the center understand the health care needs of its users. All information will be used only for statistical purposes. All answers will be held confidential. The interviews usually take about an hour and a half

In a few days, a staff member from our center will call you to answer any questions you may have, and to arrange for an interviewer to call you. In the meantime, if you have any questions about the study, please call us at the following telephone number: _____

Sincerely,

This study is authorized by Congress in Section 241 of the US Public Health Service Act (42 USC 238j).

EXHIBIT II..4.c

GRANTEE LETTERHEAD

INTRODUCTORY LETTER FOR PARENT OF SAMPLE CHILD UNDER 13 YEARS OF AGE

Dear _____

I am writing to ask you to take part in a study of people who use Community Health Centers. The study is being conducted by the U.S. Public Health Service, which supports Community Health Centers. Our center, like others across the country, provides health care to people with rnany different health care needs. Learning about those needs, and the ways that people use health centers is very important for planning the future of the health center program.

This study is being conducted in many health centers across the country. Your child has been randomly selected to represent the users of our center in this study. We are asking you, as _____'s parent to complete an interview at the community health center with a specially trained interviewer. Questions will be asked about your child's health and health care needs. The interviewer will also ask about all of the places your child goes for health care, and opinions on the care received. For completing the interview you will receive $20 to cover any expenses.

Participation in this study is voluntary. If you decide not to participate, it will have no effect on the services received at our center. However, your cooperation is very important for making the study a success and for helping the center understand the health care needs of its users. All information will be used only for statistical purposes. All answers will be held confidential. The interviews usually take about an hour and a half

In a few days, a staff member from our center will call you to answer any questions you may have, and to arrange for an interviewer to call you. In the meantime, if you have any questions about the study, please call us at the following telephone number: _____

Sincerely,

This study is authorized by Congress in Section 241 of the US Public Health Service Act (42 USC 238j).

In designing the survey it was assumed that 80 percent of the interviews would be conducted at the CHC. This was based on the assumption that the sample members were active patients and that the CHC was at a convenient location.

As it turned out, only 38 percent of the interviews were completed at the CHCs. An additional 36 percent were completed in respondent's homes or other locations. The planning assumption was that only 5 percent of the in person interviews would be completed outside of the CHC. The remainder, 26 percent, were completed by telephone; the plan had assumed 15 percent would be completed by telephone. Whenever possible, a set of response category handcards was sent to the respondent in advance of the telephone interview. If the respondent did not have the handcards at the time of the telephone interview, the interviewer read aloud the answer categories displayed on the handcards.

Language. The early collection of grantee level information indicated that, for 19 of the 48 participating grantees, more than 20 percent of their patients were non-English speaking. The most prevalent language other than English was Spanish, which included Mexican, Puerto Rican, and Cuban regionalisms. The second most prevalent language was Chinese (Mandarin), but Vietnamese and Toisanese were also represented. Of the 52 interviewers hired for the study (one resigned before training), nearly 40 percent were bilingual. This is a high percentage of bilingual interviewers to be hired for any one study.

Table II. 7 shows the distribution of completed interviews by language. Slightly more than 8 1 percent were completed in English; about 16 percent were completed in Spanish, with most respondents of Mexican descent; only two percent were completed in Chinese, but these would not have been completed without the translation of the instrument. The remainder of the non-English interviews were carried out on the English version of the CAP1 instrument with assistance from interpreters. Twenty-seven interviews made use of an interpreter. Therefore, most of the language needs were accommodated by bilingual interviewers.

TABLE II.7

COMPLETED INTERVIEWS BY LANGUAGE

Language	Number	Percent
English	1,572	81.4
Spanish	311	16.1
Chinese	39	2.0
Other (French, Vietnamese, Other)	10	0.5
	1,932	**100.0**

Immunization Data. Information on immunizations for children under age six was collected at the time of the interview whenever possible. When the appointment was set up, the adult respondent was asked to bring the child's immunization booklet to the interview. An adult who failed to bring it was subsequently telephoned in order to obtain the information in the booklet. If the adult did not have such a booklet, the information was collected from the medical record at the CHC with consent from the parent. A global question on whether the child's immunizations were up-to-date was asked in either case. Seventy percent were completed from the parent record at the time of the interview and 30 percent from the clinic record (or a follow-up call to the parent).

b. Coding of Medical Conditions

All the detailed information on the user questionnaire about conditions, injuries, and impairments needed to be systematically coded. Facilitating comparisons of this database with that of the National Health Interview Survey (NHIS) required the use of the NHIS medical coding system, a complex compilation of codes based on the ICD9 coding system and decision rules that have been developed through years of experience. The system involves the following two steps:

- First, the coder determines if a condition is chronic (i.e., the time of onset is greater than three months) or acute (i.e., the time of onset occurred within the last three months). Even these rules have specified exceptions.
- Second, the coder reviews all relevant information on the questionnaire to determine the "best" possible code. This includes consideration of the kind of disease and the body part(s) affected.

The NHIS system has rules to address problems related to unclear respondent reports of conditions; use of dual classifications; selection of one code when more than one is possible; use of separate coding when conditions do not combine; and distinguishing of symptoms from actual diseases.

Training an NHIS medical coder requires four weeks in the classroom and substantial monitoring afterwards, so we engaged two experienced NHIS medical coders who had recently retired. These coders were to follow NHIS coding protocols for the medical coding for this survey.

Because the User Survey questionnaire was computerized (CAPI), a hard-copy version of the relevant information was produced for review by the medical coders. A worksheet (Exhibit II.5), which was created and reviewed with NHIS staffbefore it was implemented, includes information from 23 questions of relevance to medical coding. The medical coders entered all relevant NHIS codes on the worksheet itself, and this information was then entered into the database.

The process followed for medical coding consisted of four major steps. First, before beginning the official coding, the two coders each coded the same 30 cases and then compared the results. All codes agreed except for one minor difference, which was resolved after discussions. A senior NHIS staff person also reviewed and approved this work. Second, the coders occasionally blind coded some of the same cases as a check. Third, all cases with questions were flagged and discussed with a senior staff member at MPR. Fourth, about three percent of the cases required checking additional information not on the worksheet. The coders made a note on the condition page in question and the possible choices of codes, and these cases were then reviewed against the CAP1 questionnaire with complete information at MPR before a final code was assigned.

Although it is not practical to describe fully the NHIS coding system, a few comments should be made. NHIS does not use certain ICD9 codes and has also developed new, unique codes as part of the NHIS system. Table II.8 lists both the ICD9 codes not used and the unique NHIS codes.

Most medical codes are numeric, with two exceptions: X-codes and E-codes. An "X" in the first digit of the code identifies the condition as is an impairment, with the next two digits indicating the type and site of the impairment, and the fourth digit indicating the cause. X-codes are always a type of chronic

EXHIBIT II.5

USER AND VISIT SURVEY MEDICAL CODING WORKSHEET

ID Number:|__|__|__|__|__|__|__|__| Respondent Age: |__|__| Sex: |__| Condition #: |__|__| ICD9 |__|__|__|__| |__|__|__|__| |__|__|__|__|

Name as Condition #:

RESP Called Cond.:

Condition was Reported in: Condition List

Reported in Limit as:

(3b) Dr. Called Cond.

Coded As:

Name of operation:

(4) Tumor/Cyst/Growth

(d3c) Resulted from Accid/Inj:

(3c) Cause of Condition: Specify

(5) When first noticed/occurred:

(3d) How accident happened:

(3e) What kind ofcondition:

(3f) How allergy/stroke now affect:

(3g) Part of body affected:

(3h) Part of body affected by infection/soreness:

(7a) At time of accident part of body was hurt:

Kind of injury:

(7b) What part of body affected now:

How affected now:

TABLE II.8

LISTING OF ICD-9 CODES NOT USED **AND** NHIS UNIQUE CODES

I. NHIS DOES NOT USE THE FOLLOWING ICD CODES:

0091	319	391	740	7616	9086
0093	326	410	741	7680-7681	9090-9091
137	3313	411	7420	7796	9093
138	3314	412	7421	7815	9099
139	342	4296	7423	7831	9 1 O-9 19 (4th digits)
230-234	343	436	743	798	920-924 (4th digits)
235-238	3440	438	7440-7443	800-829 (4th digits)	925-929 (4th digits)
2681	3442	5200	7445	830-839 (4th digits)	930-939 (4th digits)
3060	3443	5201	7448-7449	840-848 (4th digits)	940-949 (4th digits)
3066	3444	5202	749	850-854 (4th digits)	950-957 (4th digits)
3067	3445	5205	7500-7501	860-869 (4th digits)	958
3068	3446	5216	7530	870-897 (4th digits)	959 (4th digits)
3069	3448	524	7542-7548	8713	960-969
3150	3449	7186	755	885897	970-977.8
3151	3535	7282	7560-7563	885-887	978-979
3152	369	728 4	7580	895-897	980-987
317	3882	7286	7597	900-904 (4th digits)	9890-9894
318	389	734-738	7599	9056-9059	9896-9898

II. NEW UNIQUE CODES DEVELOPED BY NHIS

0 19	Tuberculosis stated to be arrested or inactive[a]	3 167	Organ of Special Sense"
3 160	Musculoskeletal"	3168	Other[a]
3 16 1	Respiratory[a]	3 169	Unspecified"
3 162	Cardiovascular"	399	Rheumatic fever inactive (old) (History)
3163	Skin"	4777	Allergic with multiple causes
3 164	Gastrointestinal"	6932	Skin allergies with multiple causes
3 165	Genitourinary"	7995	Observations with no condition found
3 166	Endocrine"		

"Physical factors associated with diseases classified elsewhere. These codes are not used alone; an additional code is used to identify physical condition.

condition. E-codes are a supplementary classification for external causes of injuries and poisonings. E-codes usually indicate a one-time occurrence. The place of occurrence of the accident is indicated by one of the digits. A NHIS rule requires coding symptoms and all ill-defined statements of "trouble" (for example, stomach ***trouble*** instead of ***upset)*** whether listed as a trouble, disease, attack, or other categorization. A complete description of the system can be found in the document entitled ***"Public Use Data Tape Documentation, Part Three: Medical Coding Manual and Short Index,"*** U.S. Department of Health and Human Services, 1988.

There were about 18 persons whose conditions and corresponding codes needed to be hand-edited. These were cases where the respondents changed their mind during the interview about the conditions they had after the grid of conditions had been established by the CAP1 program, either adding or deleting a condition. These were resolved and recoded on a case-by-case basis. We then checked that each condition had a corresponding code or codes. Once the condition codes were all entered, verified, and updated, the condition code file (which contained one record per condition) was unduplicated; that is, for cases where the coder instructed us to combine two or more conditions together under one code (or set of codes), the duplicate records were removed from the file. Finally, a variable indicating the question that generated the condition (Condition List or type of limitation) was added to the condition file.

One medical coding problem that arose unexpectedly had to do with the Condition List. Whenever a respondent claimed to have had one of the conditions listed, a Condition Page was generated for that condition. The condition itself was automatically listed verbatim as it appeared on the Condition List by the CAP1 program; for example, "Permanent stiffness or any deformity of the foot, leg, fingers, arm, or back." While the specific ailment was ultimately recorded as part of the Condition Page questions, the condition coders received a coding sheet labelled with the general statement, "Permanent stiffness or any deformity of the foot, leg, fingers, arm, or back." We later discovered that the coders were trained to code such conditions as if **the** person had problems with his foot, leg, fingers, arm, ***and*** back; that is, the coders

were trained to code all conditions the respondent mentioned in the verbatim condition, regardless of the responses to the subsequent questions. The reason is, on the hard copy NHIS, interviewers probed and only wrote down the relevant part of the phrase from any Condition List entries, whereas the CAP1 program took the entire phrase verbatim. Once we realized this was happening, the coders went through the cases again and only coded the relevant part of the condition phrase.

Despite the procedures set up to ensure the accuracy of the coding process, when the coded cases were reviewed by MPR staff, and when the blind double-coded cases were reviewed, numerous questions were raised about the accuracy of the recorded codes. After discussions with the coders, we decided to have the work reviewed by coders currently working for NHIS. That review was to indicate codes where there was agreement; codes that could be coded differently, but where the preferred code was not clearly indicated; and codes that should be changed. A sample of cases in the latter two categories was sent to NCHS for review and adjudication. Out of about 50 cases sent, NCHS classified 35 percent as the original coder being right; 38 percent where the current NHIS coder was right; and 27 percent where neither was right. We opted to use the decision of the current NHIS coder in cases where there was a discrepancy.

c. Other coding and editing

In addition to coding medical conditions, we coded three other verbatim responses in the user survey: occupation, industry, and surgical procedures while hospitalized. Occupation was coded according to the *Standard Occupational Classification Manual*[2], based on the verbatim responses to the occupation question ("What kind of work were you doing?') and the question about duties on the job ("What were your most important activities or duties at that job?'). Industry was coded according to the *Standard Industrial Classification Manual*[3], based on the verbatim responses to the industry question ("What kind

---

[2]U.S. Department of Commerce, Office of Federal Statistical Policy and Standards, 1980.

[3]Executive Office of the President, Office of Management and Budget, 1987.

of business or industry is this?' and the employer question ("For whom did you work?'). Both occupation and industry were coded to the two-digit level of specificity.

In the Source of Care question, respondents were asked about all places they went to for health care in the past 12 months. For each place, they were asked a series of questions. If the place was a hospital they stayed in overnight, they were asked, "Did you have any kind of surgery or operation during (any of) the stay(s) in the hospital?' If they responded that they had, they were asked, "What was the name of the surgery or operation? Were there any others?" These verbatim responses were coded according to the *Physicians ' Current Procedural Terminology* manual[4]. The operation or surgical procedure was coded to the five-digit level of specificity for up to five procedures per person.

Most of the editing for the user survey was built into the CAP1 program in the form of fixed response categories, valid ranges, and verify screens. The verify screens signaled to the interviewer that a highly improbable or contradictory response had just been entered, at which point the interviewer could check whether this was due to a keying error on his or her part, or could ask the respondent to verify the response (at which point a correction could be made to that or a previous response, if necessary). There were a few variables that appeared to have out-of-range values (P17AU, F4BP, and H2P). It would have been difficult to build in a verify screen for them because the responses for these three questions could have been reported in one of a few different units. It was not until the responses were calibrated to a single unit that the improbable values were apparent. These were edited and saved under new variable names: P17AUED, F4BPED, and H2PED. Some other minor editing took place for: the immunization section (ensuring chronological dates and comparing dates of immunizations with birthdate); and checking for consistency in reports of military status (Condition Page report of injury while in military versus report of veteran status in Demographic section).

[4]American Medical Association, 1994.

d. Recommendations

NCHS no longer codes medical conditions in the NHIS. If a survey of CHCs is conducted in the future, this practice should be adopted for that survey.

3. Response Patterns

a. Response rate

The final statuses for all selected cases can be found in Table II.9. Of the 2,929 medical users selected 149 were frame ineligible, leaving **2,780** available for release to the field. Among these, 1,924 completed the entire interview and another 8 completed at least half of the interview and were judged to be usable observations.[5] Of the 1,932 completes, 735 (38 percent) were completed at the community health center, 496 (26 percent) were completed over the telephone, and 701 (36 percent) were completed in person at a location other than the community health center (usually the user's home). Among the 1,932 completes, 1,313 (68 percent) were self-respondents and 619 (32 percent) were adults responding as a proxy for either a sampled child under age 13 or an adult unable to respond for himself or herself. Among the self-respondents, 17 interviews were conducted using an interpreter. Out of the 619 proxy interviews, 10 required an interpreter.

There were 239 eligible cases who did not respond to the survey and were known to be eligible. These eligible nonrespondents were classified as follows: refusals (n=113), maximum contacts (n=48), other nonresponse (n=48), temporarily out of area (n=12), breakoffs (n=8), language barrier (n=5), too ill (n=4), and no proxy available (n=1). "Refusals" are cases where the selected person refused to participate, even after several attempts were made to change the person's mind. "Maximum contacts" means that an attempt to contact the person was made at least 10 times (but not more than 30), but the

[5]How to classify partial completes was determined by whether the interview had been completed up until the Condition Page section, so that all conditions had been enumerated. Those that did not complete all sections prior to the Condition Page were considered breakoffs (nonrespondents). The remainder were considered completes.

TABLE II.9

USER SURVEY FINAL STATUSES

			Percent of Selected Sample		Percent of Completes	
Total Sample Selected	2,929		100.0%		--	
Frame Ineligible	149		5.1			
Ineligible	217		7.4		--	
Migrant farmworker		21		0.7		--
Homeless		6		0.2		
Moved out of area		123		4.2		--
Institutionalized		10		0.3		--
Deceased		26		0.9		--
Other		31		1.1		--
Nonlocatable	392		13.4			
Eligible Nonrespondent	239		8.2			
Breakoff		8		0.3		--
Final refusal		113		3.9		
Maximum contacts		48		1.6		
Language barrier		5		0.2		--
Out of area during data collection		12		0.4		
Too ill		4		0.1		
No proxy available		1		0.0		
Other		48		1.6		--
Completes	1,932		66.0		100.0	
At clinic		735		25.1		38.0
Adult respondent			454	15.5		23.5
Teen respondent			64	2.2		3.3
Child/proxy			202	6.9		10.5
Other proxy			15	0.5		0.8
Telephone		496		16.9		25.7
Adult respondent			295	10.1		15.3
Teen respondent			39	1.3		2.0
Child/proxy			143	4.9		7.4
Other proxy			19	0.6		1.0
Other location		701		23.9		36.3
Adult respondent			412	14.1		21.3
Teen respondent			49	1.7		2.5
Child/proxy			223	7.6		11.5
Other proxy			17	0.6		0.9

interviewer was never able to speak to the selected person. These are different from "nonlocatables" (see below), because it is clear to the interviewer that s/he had reached the home of the selected person and believed this person was eligible to participate. "Other nonresponse" generally means that the interviewer was unable, after several attempts, to arrange an interview before the field period ended. "Temporarily out of area" means that the person was known to have been away, during the entire period of data collection, from the service area from which he or she was sampled. "Breakoffs" are cases where the interview was interrupted after it began and the respondent refused to complete the interview at that time or at another time. "Language barrier" means that the person did not speak the language of the interviewer, and no interpreter was available. "Too ill" means the person was unable to complete the interview because of his or her health. "No proxy available" means the selected person either was a child or was an adult unable to communicate owing to a mental, physical, or developmental disability, and no proxy was available to respond to the interview.

Of the 609 remaining cases, 217 were classified as ineligible after an attempt was made to contact the user. Most of the ineligible sample members had moved away from the CHC's service area since their last visit (n=123). The remainder of the ineligible sample members had the following characteristics: the selected record did not correspond to a 1994 medical user (n=3 1); deceased (n=26); migrant farmworker (n=2 1); institutionalized (n=10); and homeless (n=6). Finally, there were 392 selected users that could not be located. It is unknown whether these users would have been eligible had they been contacted. These are discussed in more detail in the next section.

The overall response rate is generally defined as the number of eligible completes divided by the number of eligible users. However, because about 15 percent of the eligible sample could not be located, we have to estimate the number of eligible and ineligible cases among that group (Table II. 10). The estimated number of eligible users is, then, **the sum of the number of eligible respondents, the number of** eligible nonrespondents, and a fraction of the nonlocatables, that is, those with unknown eligibility status.

TABLE II. 10

USER SURVEY RESPONSE RATE

	N	Percent	Percent of Confirmed Eligibles
Total Sample Fielded	2,780	100.0	
Cases Determined Ineligible	217	**7.8**	
Non-locatable Cases	392	**14.1**	
Total Sample Confirmed Eligible	2,171	**78.1**	**100.0**
Refusals, Breakoffs, and Other Non-response	239	**8.6**	**11.0**
Completed Interviews	1,932	**69.5**	**89.0**
RESPONSE RATES			
Adjusted		**76.4**	
Unadjusted[b]		**75.4**	

Conservatively assumes that the eligibility rate among the non-locatable cases is the same as the confirmed eligibility rate among all cases fielded or 2,171 ÷ (2,171 + 217) = .909. This means that 356 of the non-locatable cases would be eligible; 1,932 ÷ (2,171 + 356) = .764. Note: It is likely that the ineligibility rate among non-locatables is greater than for the rest of the sample.

[b]This is total completes divided by all cases fielded minus only cases with confirmed ineligibility. This rate makes the unrealistic assumption that all non-locatables are eligible.

The fraction used is the eligibility rate as determined from those with known eligibility status. The eligibility rate used here is conservatively estimated at .909; that is, the number of known eligibles (2,171) divided by known eligibles plus known ineligibles (other than frame ineligibles) (2,171 + 217). This estimate is conservative in that it assumes that the nonlocatables have the same eligibility distribution as those who were located, whereas the true eligibility rate among nonlocatables is likely to be lower. (Recall that most of the cases that were confirmed ineligible had moved out of the area. It is likely that an even higher proportion of the nonlocatable cases had moved out of the area.) Using this eligibility rate probably has the effect of producing a lower overall response rate. Applying this rate to the nonlocatables yields an estimate of 356 eligibles out of 392. The overall response rate can then be calculated as 1,932/(2,171+356) = ,764, or 76.4 percent.

As can be seen in Table II. 10, response rates varied somewhat by sex, age group, urbanicity, and region. Women were more likely to respond than men, and sample members ages 18 to 40 were the least likely to respond. But the most marked differences in response rates were related to whether the CHC was located in an urban or rural area and the region in which the CHC was located. Those selected from rural CHCs were much more likely to respond than those from urban CHCs. Persons selected from CHCs located in the West were most likely to respond while those selected from a Northeast CHC were the least likely to respond.

Table II. 11 also shows various types of nonresponse by sample characteristics. Overall, among those known not to be ineligible, 4.4 percent refused to be interviewed. There was no difference in refusal rates between males and females; however, a large difference can be seen between those under and over age 40. Those in the oldest age category (41 and older) were twice as likely to refuse (7.3 percent) as those in the two younger age categories (3.0 percent and 3.9 percent). Of course, it would be the parent deciding whether to participate or refuse, in the case of those under age 18. Sample members in urban areas were

TABLE II. 11

USER SURVEY RESPONSE RATES

BY SAMPLE CHARACTERISTICS

Characteristic	Adjusted Response Rate"	Unadjusted Response Rate"	Refused[b]	Nonlocatable	Other Nonresponse[c]
Total Sample	76.4%	75.4%	4.7%	15.3%	4.6%
Males	72.6	71.2	4.7	18.9	5.2
Females	78.5	77.9	4.7	13.2	4.2
Ages 0 to 17	76.9	76.2	3.4	15.7	4.6
Ages 18 to 40	74.7	73.3	4.2	17.0	5.5
Ages 41 and older	77.6	76.7	7.3	12.5	3.5
Urban CHCs	71.8	70.7	5.2	19.0	5.1
Rural CHCs	82.0	81.1	4.2	10.8	4.0
Northeastern CHCs	69.1	67.6	5.2	24.5	2.7
Midwestern CHCs	71.4	69.8	4.8	20.4	5.0
Southern CHCs	75.6	74.7	4.6	14.4	6.3
Western CHCs	86.1	85.6	4.5	6.7	3.3

"See Table II. 10 for definitions and derivation of response rates.

[b]Refused = refusals + breakoff.

"Other nonresponse = Maximum contacts, language, out of area, too ill, no proxy, and other.

somewhat more likely to refuse than those in rural areas (4.8 versus 3.9 percent), and those living in the Northeast were more likely than average to refuse (5.2 percent) while those in the West were less likely than average to refuse (3.6 percent).

A much larger reason for nonresponse was the inability to locate sample members. Overall, the rate of nonlocatables was 15.3 percent. The problem was larger for males (18.9 percent) than females (13.2 percent). There was not much of a difference in the nonlocatable rate by age group. Geographical characteristics of sample members were a major factor in the rate of nonlocatables. Those in urban CHCs had a nonlocatable rate of 19.0 percent, whereas those in rural areas had a comparatively lower rate of 10.8 percent. Sample members in the Northeastern and Midwestern parts of the country had nonlocatable rates of over 20 percent, whereas those in the West had a nonlocatable rate of only 6.7 percent.

Appendix 4 contains refusal and nonlocatable rates by grantee. Ten grantees had no refusals at all. Among the remaining 38 grantees, the refusal rates ranged from as low as 1.5 percent to as high as 17.5 percent. Only one grantee had no cases that were nonlocatable. Among the remainder, the nonlocatable rates ranged from 1.6 percent to 68.3 percent per grantee. The highest nonlocatable rate was in a large central city where apartment numbers were not included in the contact information provided by some sites.

b. Item nonresponse and other issues

The CASES program offered the categories "don't know" and "refused' for virtually all questions in the interview. Except for text entries and intentional skips, the program does not allow for any questions to be left unanswered. (In the case of the eight breakoffs that were considered to be complete, the unanswered questions were filled with "don't know" responses according to the logical skips.)

Table II. 12 shows the number of responses for a set of key variables. Among the Condition List variables (beginning with the letter "C"), there were virtually no refusals; however, several conditions resulted in a number of "don't know" responses: anemia (n=16), arthritis (n=14), congenital heart disease (n=14), coronary heart disease (n=12), hypertension (n=11), and other circulatory problems (n=11). All

TABLE II. 12

ITEM NONRESPONSE FOR KEY VARIABLES

Variable		Response	Don't Know	Refusal	Skip
C1a	Stiffness	1930	2	0	0
C1b	Arthritis	1918	14	0	0
C1c	Neck, back, spine	1931	1	0	0
C1d	Dermatitis	1932	0	0	0
C2a	Deafness	1925	7	0	0
C2b	Trouble hearing	1927	5	0	0
C2c	Blindness	1931	1	0	0
C2d	Trouble seeing	1925	7	0	0
C2e	(Cerebral) Palsy	1924	8	0	0
C2g	Accident/Injury	1926	6	0	0
C5a	Arteriosclerosis	1171	4	0	757
C5b	Congenital heart disease	1918	14	0	0
C5c	Coronary heart disease	1163	12	0	757
C5d	Hypertension	1921	11	0	0
C5e	Stroke	1929	3	0	0
C5f	Angina	1168	7	0	757
C5g	Heart attack	1173	2	0	757
C5h	Cancer	1930	2	0	0
C5i	Lead poisoning	605	3	0	1324
C5j	Other heart trouble	1929	3	0	0
C5k	Other circulatory	1921	11	0	0
C3a	Gallbladder	1171	4	0	757
C3b	Cirrhosis	1172	3	0	757
C3c	Hepatitis	1926	6	0	0
C3d	Ulcer	1170	5	0	757

TABLE II. 12 *(continued)*

Variable		Response	Don't Know	Refusal	Skip
C3e	Indigestion	1930	2	0	0
C3f	Other stomach	1930	2	0	0
C3g	Other digestive	1929	3	0	0
C4a	Goiter/Thyroid	1925	7	0	0
C4b	Diabetes	1924	8	0	0
C4c	Anemia	1916	16	0	0
C4d	Epilepsy/Seizures	1931	1	0	0
C4e	Kidney	1927	5	0	0
C4f	Bladder	1927	5	0	0
C4g	Genital	1929	3	0	0
C4h	Prostate	335	1	0	1596
C4i	Breast cancer	839	0	0	1093
C4j	Uterus/Ovaries	833	5	1	1093
C4k	Other female	1254	5	0	673
C6a	Bronchitis	1924	8	0	0
C6b	Asthma	1925	7	0	0
C6c	Emphysema	1171	4	0	757
C6d	Tuberculosis	1927	5	0	0
C6e	Tonsillitis	1929	3	0	0
C6f	Work-related respiratory	1144	4	0	784
C6g	Other respiratory	1928	4	0	0
E2a	Education	1652	2	1	277
E3a	Hispanic	1926	5	1	0
E4a	Race	1920	11	1	0
E4b	Best race (if more than 1)	146	7	3	1776
E5a	Working	1173	0	2	757
E8	Marital status	1321	2	1	608

TABLE II. 12 (continued)

Variable		Response	Don't Know	Refusal	skip
E9b	Family income	1158	118	17	639
El Oa	Country of birth	1930	0	2	0
MO	CHC usual source	1926	3	3	0
M1a	If not, have usual source	322	2	3	1605
M1b	If more than one, one most often	11	0	0	1921
M5ba through M5bm	Center convenient No other place	1599	5	0	328
		1599	5	0	328
M6ba	Type of place 1	1282	3	0	647
M6bb	Type of place 2	617	9	0	1306
M6bc	Type of place 3	199	2	0	1731
M6bd	Type of place 4	62	0	0	1870
M6be	Type of place 5	23	0	0	1909
P1	Hypertension	1556	3	0	373
P10	Cholesterol	916	25	0	991
P12	Diabetes	1556	3	0	373
P25a	Asthma	1815	10	0	107
U1a	Bathing	1172	2	1	757
U1b	Dressing	1174	0	1	757
U1c	Eating	1174	0	1	757
U1d	Transfer	1174	0	1	757
U1e	Toileting	1174	0	1	757
U1f	Getting around house	1174	0	1	757
X1	Medicare	1915	9	8	0
x2	Medicaid	1903	21	8	0
x3	Other public insurance	1905	19	8	0
X5	Military insurance	1916	8	8	0

TABLE II. 12 (continued)

Variable	Response	Don't Know	Refusal	Skip
X6 Private insurance	1898	26	8	0
Z1a Ever had pap	826	6	7	1093
Z3a Heard of mammogram	555	3	6	1368
Z3b Ever had mammogram	489	0	0	1443
Z 1 bmon Most recent pap	745	40	0	1147
Z3 dyr Most recent mammogram	341	18	0	1573

other conditions had fewer than ten "don't know" responses. Similarly, in the Chronic Disease Followup section (variables beginning with the letter "P"), there were no refusals; however, there were 25 persons who responded "don't know" to the question about whether they were ever told that they had high cholesterol, and 10 persons who responded "don't know' to a similar question about asthma. With respect to limitations in activities of daily living (beginning with the letter "U"), only one person refused, and two persons responded "don't know" to the question about bathing.

Among the sociodemographic variables (beginning with "E"), there were 17 refusals on the family income question, and 118 "don't know" responses. There were only a handful of "don't know" and "refused" responses for the other demographic variables. Only eight persons refused to answer the insurance questions (beginning with "X"), but a number of people responded "don't know" to questions about private insurance (n=26), Medicaid (n=21), and other public insurance (n=19). Only three persons refused to answer the question about having a usual source of care (beginning with "M"), and a handful of persons responded "don't know" to each of the questions about the usual source of care and other sources of care.

In the Cancer Screening section (variables beginning with "Z"), only six women refused to answer the question about whether they had ever heard of a mammogram, and three did not know the answer. Eighteen women did not know when they had last had a mammogram. Seven women refused to answer the question about whether they had ever had a Pap smear, and six did not know the answer. Forty women did not know when they last had a Pap smear.

Any missing (blank or null) value for a questionnaire item (other than text entries) is the result of a logical skip, meaning that the question was intentionally not asked because it does not apply to the respondent or because the response is obvious from a prior response. When calculating estimates or constructing new variables, it may be desirable to set some of the logically skipped variables to their implicit values. In addition, it will generally be necessary to convert the "don't know" and "refused"

responses to a missing value (in SAS, that value would be a dot for numeric variables) and perhaps statistically impute a non-missing value to such cases using a procedure such as hot-deck imputation. For descriptive purposes, it may be desirable to leave the "don't know" and "refused" categories intact.

As described earlier, for all sampled children under age six, the immunization section of the questionnaire was administered. If the child's parent or guardian had not brought the child's personal immunization records to the interview, the parent/guardian was asked for permission to look at the child's medical records for this information, and some summary questions were asked at the time of the interview. In cases where the personal immunization record was available at home, the detailed immunization information was later obtained from the parent/guardian later over the telephone. In all other cases, an attempt was made to obtain the detailed information from the CHC chart. In 23 cases, a CHC staff member looked at the child's chart and found no immunization information listed. In that case, the relevant responses in the immunization section were coded as -3 and IMMUSTAT was coded as 3. In 11 cases, the CHC was unable to locate the child's chart promptly. In that case, the relevant responses in the immunization section were coded as -4 and IMMUSTAT was coded as 4.

There are other types of item nonresponse for coded and constructed variables. When the condition coders were unable to assign an ICD9 code to a particular condition because of insufficient or contradictory information, a code of 7998 or 7999 was assigned. When the coder of industry (E7_C) and occupation (E7_D) was unable to assign a standard industrial classification (SIC) or standard occupational classification (SOC) code from what was reported by the respondent, a code of "U" was assigned. Similarly, when the coder of surgical procedures and operations (M7Hi_CD, i=A,B,C,D,E) was unable to assign a CPT code, a code of "U" was assigned. For each questionnaire item that had multiple units in which the response could have been given, a constructed variable that calibrated all responses to a single unit was created. If either the amount variable had a "don't know" or "refused" value or the unit variable was otherwise logically skipped, then the constructed calibrated variable was assigned a missing value.

c. Nonlocatables

Among those sample members found to be eligible for the user survey, 89 percent completed the interview (see above), so refusals and other types of nonresponse did not have a large effect on the number of completes. In addition to the large number of persons who were found to be ineligible because they had moved out of the service area, the other major source of problems in obtaining completed interviews was the inability to locate a large number of selected users (n=392). Because changing residence is more common among lower-income populations, particularly in urban areas, it was expected that a large proportion of the sample members would have moved since they were last seen at the CHC; however, it was not anticipated that the CHC would provide locating information that was often out of date, incomplete, or insufficient.

One-third of the CHCs had at least ten nonlocatable cases each. The number of nonlocatables ranged from 0 at one CHC to 28 at another. The CHC that had 28 nonlocatables was in an urban location, where many users lived in high-rise apartment buildings or did not have working telephones. Because the CHC was not able to provide apartment numbers, it was difficult if not impossible to locate many of the sample members. In CHCs located in areas with large numbers of illegal aliens, such as those near the Mexican border, some addresses found in the charts turned out to be nonexistent. It should be noted, however, that the nonlocatable problem was less severe in the Western states.

Interviewers worked both independently and with CHC staff to try to locate all selected users. Once an interviewer determined that a sample member was not at the indicated address and that the CHC did not have more up-to-date information, the interviewer tried to locate the respondent using a number of techniques, including checking with directory assistance; inquiring at the local post office about a forwarding address; contacting neighbors, local shopkeepers, and the local mail carrier; and calling other households with the same surname (except in cases with a common surname).

We found that the field interviewer's efforts to locate users were more effective than a central office search using automated databases. While most of the nonlocatable cases probably had moved out of the CHC's service area, in the absence of any evidence to that effect, these cases were considered to have an unknown eligibility status. As discussed above, in calculating the response rate, we assumed that about 91 percent were eligible.

While we had anticipated a certain degree of nonlocatable cases in this population, the magnitude of the problem was higher than expected and worse than MPR has experienced in other studies with similar populations. For example, in the Teenage Parent Demonstration, less than one percent of cases were nonlocatable for a two-year followup and about nine percent were nonlocatable after five years. In these other studies, we were generally able to use state databases or databases of program participants (such as Medicaid or WIC). In this study we were relying on databases maintained by individual health centers. In addition, many patients are described by CHC staff as "one-time-onlies"--patients who visited the center only once. Such patients are much more difficult to trace.

d. Other Fielding Issues

This section will address the mode of the completed interviews, the time required to complete the interviews, interviewer attrition, interviewer travel expenses, and CHC cooperation with the user survey once it was fielded.

As stated previously, seventy-four percent of the interviews were completed in person, usually at the clinic or in the respondent's home, but sometimes at another location. The other 26 percent, a higher-than-predicted figure, were completed by telephone. More were completed outside the clinic than anticipated because of chronic no-shows at the clinic, the preference of the respondent, or the fact that the interviewer was located closer to the respondent's home than to the clinic.

Table II.13 indicates the distribution of completed interviews by month. The original plan was to complete the user interviews by August 1995, but the actual field period ran from April 1995 to February

TABLE II. 13

COMPLETED INTERVIEWS BY MONTH

Month	Number of Completes	Percent	Cumulative Percent
April 1995	7	0.4	0.4
May 1995	368	19.0	19.4
June 1995	524	27.1	46.5
July 1995	427	22.1	68.6
August 1995	269	13.9	82.5
September 1995	167	8.6	91.1
October 1995	136	7.0	98.1
November 1995	21	1.1	99.2
December 1995	7	0.4	99.6
January 1996	1	0.1	99.7
February 1996	5	0.3	100.0
	1,932	**100.0**	

1996. Approximately 20 percent of the completed interviews were done between September and January. The reasons for the extended field period include the following:

- Some grantees experienced difficulty constructing the user sampling frames, thus slowing the operations which depended on them.
- CHC staff often turned over, which delayed or temporarily suspended survey operations.
- Some of the grantees, particularly those serving large numbers of transient CHC users, had large numbers of people who were difficult to locate.

Each grantee had unique requirements for the survey, and each had different systems and levels of personnel to support the sampling process. Therefore, each grantee had to be managed as a separate 'entity with unique needs and a unique schedule. As a result, many interviewers could not begin for weeks or sometimes even months after the training session. This required the creative use of their time and some retraining as well. Interviewers helped to expedite the sampling and study start-up whenever possible by offering to assist the clinic staff Other interviewers were faced with starts and stops. While interviewers for this study were particularly committed, there was a need for encouragement to sustain morale through the unexpected delays.

Under the circumstances, interviewer attrition was at a reasonable level. One interviewer resigned before training, one immediately after training. In six instances, interviewers had to be sent to another area in order to complete the interviews. The need for traveling interviewers resulted from a combination of factors: the extended field period; the need for bilingual interviewers in some places where it was not expected; and, in a few cases, illness or other interviewer limitations. The problems due to interviewer attrition were minimized by having interviewers at one location complete interviews by telephone at another location.

The grantees were highly cooperative. Once a grantee agreed to participate, problems were usually related to the technical difficulties of providing information for sampling. Many of the grantees used their

files to locate sampled respondents and provide other information. Grantees provided space at the clinics for the interviewing.

4. Weighting Methodology and Design Effects

This section outlines the strategy used to create sampling and analysis weights for the user survey. As used here, a *sampling weight* is the reciprocal of the probability of selection for each sampled unit. An *analysis weight* is a sampling weight that has been adjusted for nonresponse, and possibly poststratified to known population totals and/or trimmed to mitigate the effect of outlier weights. Cluster and grantee weight components are discussed first below. Following that is a discussion of the selection of sites and users, and weighting class adjustments for users.

a. Grantee-Level Sampling Weight

The first two sampling steps in both the User Survey and Visit Survey were the selection of (1) grantee clusters within strata and (2) grantees within grantee clusters. For larger grantees, there may have been only one grantee cluster per stratum or only one grantee per grantee cluster, in which case stage (1) or stage (2) reflects a certainty selection. See Appendix 5 for definitions of grantee clusters and strata and for a description of the grantee selection process. Grantee clusters and grantees were selected with probability proportional to size (PPS).

To maintain simplicity in the notation for the formulas that follow, each subscript is meant to imply subscripts of prior stages of selection; that is, the subscript for cluster C_i implies that it is within stratum h; the subscript for grantee G_j implies that it is within cluster C_i and stratum h; site S_k implies grantee G_j, cluster C_i, and stratum h; user U_q implies (site S_k,) grantee G_j, cluster C_i, and stratum h; encounter V_r implies grantee G_j, cluster C_i, and stratum h.

The probability of selection for cluster C_i in stratum ***h*** is quantified as:

$$(1)\quad P(C_i) = \frac{m_h C\hat{M}OS_i}{\sum_{z=1}^{M_h} C\hat{M}OS_z}$$

where:

m_h is the number of grantee clusters selected from stratum h

M_h is the total number of grantee clusters in stratum h

$C\hat{M}OS_i$ is the estimated measure of size (number of 1993 medical users from January 1994 BCRR) of grantee cluster C_i

The sampling weight associated with cluster C_i in stratum h is:

$$(2)\quad W_S(C_i) = \frac{1}{P(C_i)}$$

The conditional probability of selection for grantee G_j given that cluster C_i is selected can be quantified as

$$(3)\quad P(G_j|C_i) = \frac{G\hat{M}OS_j}{C\hat{M}OS_i}$$

where $G\hat{M}OS_j$ is the estimated measure of size of grantee G_j. Note that within each cluster C_i the $G\hat{M}OS_j$ size measures total to $C\hat{M}OS_i$.

The unconditional probability of selection for grantee G_j within cluster C, is the product of the cluster C, selection probability and the G_j conditional selection probability, or:

$$(4)\quad P(G_j) = P(C_i)\ P(G_j|C_i)$$

The grantee-level sampling weight for grantee G, is obtained as the reciprocal of this selection probability, or:

$$(5)\ W_S(G_j) = \frac{1}{P(G_j)}$$

Note that this sampling weight is equal to 1 for the grantee which was designated as a certainty selection, because $m_h = M_h = 1$ and $C\hat{M}OS_1 = G\hat{M}OS_j$.

None of the selected grantees are known to have had multiple chances of selection (that is, we believe that each grantee was represented by only one record on each stratum's grantee frame), so no multiplicity adjustment was made to the grantee-level sampling weight.

For the User Survey, the next step in the sampling process was the selection of sites (that is, locations or facilities) within grantee. This stage did not apply for grantees with only one site, or for two-site grantees with dissimilar sites, where both sites were included in the sample. Site were selected with probability proportional to size within grantees.

b. Site-Level Sampling Weight

The conditional probability of selection for site S_k given that grantee G_j is selected can be quantified as:

$$(6)\ P(S_k \mid G_j) = \frac{s_j S\tilde{M}OS_k}{G\tilde{M}OS_j}$$

where:

s_j is the number of sites selected from grantee G_j

$S\tilde{M}OS_k$ is the measure of size of site S_k within grantee G_j, as stated in the grantee interview

$G\tilde{M}OS_j$ is the measure of size of grantee G_j, as stated in the grantee interview.

Note that this probability of selection can be more than one if s_j is greater than one; that is, if more than one site was selected. There were eight grantees with more than one site selected. A decision was made not to make any sites in these grantees having a probability of selection greater than one into certainty

selections. Instead, if such a site was selected twice, then twice the number of users would have been selected from the site. Four sites selected from these grantees had a probability of selection greater than one. While we allowed for multiple selections for these four sites, none ended up being selected more than once.

The unconditional probability of selection for site S_k from grantee G_j is:

$$(7) \quad P(S_k) = P(G_j)\ P(S_k|G_j)$$

In the case of grantees for which no site subsampling took place (ten grantees with only one site and six grantees with two dissimilar sites each), we set $s_j = 1$ and $\tilde{SMOS}_k = \tilde{GMOS}_j$; hence, the probability of selection of the "selected site" is equal to the probability of selection of the grantee.

Two grantees could not provide a count of 1994 users for two sites that opened late in 1994, nor could they have produced a user frame had these sites been selected. These sites without user frames were excluded from the possibility of selection. For the two grantees containing an excluded site, $\tilde{GMOS}_j$ includes the counts of medical users only from the sites included in the sampling frame.

The sampling weight for site S_k is:

$$(8) \quad W_S(S_k) = \frac{1}{P(S_k)}$$

None of the selected sites are known to have had multiple chances of selection (that is, we believe each site was represented by only one record on each grantee's site frame), so no multiplicity adjustment was made to the site-level sampling weight.

c. User-Level Sampling Weight

The fourth and final step in the sampling process for the User Survey was the selection of medical users within site or grantee. Medical user records were selected with equal probability at the last stage.

Replicate samples of users were selected at the same time as the original samples, in anticipation of having to add sample at a later time. We allowed for this possibility because, at the time the initial samples were drawn the number to be selected from each grantee was based on (among other things) the sum over all estimated frame sizes, a predicted eligibility rate, and estimated contact and response rates. Random selections from these replicates were released as needed during the field period to supplement the sample. It should be noted that a few grantees were unable to provide contact information for these random selections from the user sample replicates in a timely manner (3 5 cases in 5 grantees). Because there was never an attempt to locate or interview these cases, they were considered not to be released. For weighting purposes, it is as if they were never selected as additional sample.

For grantees where site subsampling took place, the conditional probability of selection for user record U_q given that site S_k is selected can be quantified as:

$$(9)\quad P(U_q|S_k) = \frac{u_k}{SMOS_k}$$

where:

u_k is the number of user records selected and released for site S_k

$SMOS_k$ is the actual number of user records on the site S_k frame

Note that u_k includes both the original and replicate samples selected and excludes 20 cases not released in three grantees (due to their not providing contact information in a timely manner). Moreover, u_k and $SMOS_k$ (and u_j and $GMOS_j$ discussed below) contain sampled user records known to be ineligible prior to contact as well as others later found to be ineligible. Those known to be ineligible from frame information (before contact) include 149 cases in two grantees (non-medical users and non-l 994 users). The grantees could not remove these cases from the frame prior to sample selection (except manually, which would have been an onerous task).

The unconditional probability of selection for user record U_q from site S_k is:

$$(10)\quad P(U_q) = P(S_k)\ P(U_q|S_k)$$

For grantees where no site subsampling took place, the conditional probability of selection for user record U_q given that grantee G_j is selected can be quantified as:

$$(11)\quad P(U_q|G_j) = \frac{u_j}{GMOS_j}$$

where:

u_j is the number of user records selected and released from grantee G_j

$GMOS_j$ is the actual number of user records on the grantee G_j frame

Note that u_j includes both the original and replicate samples and excludes 15 cases not released in two grantees (due to their not providing contact information in a timely manner).

The unconditional probability of selection for user record U_q from grantee G_j (with no site subsampling) is:

$$(12)\quad P(U_q) = P(G_j)\ P(U_q|G_j)$$

Thus, the sampling weight for a user record U_q is:

$$(13)\quad W_S(U_q) = \frac{1}{P(U_q)}$$

Note that this sampling weight is user *record* level as opposed to *user* level At this point, the sampling weight has a minimum value of 548.3 and a maximum value of 2096.4. The sum of these sampling weights is 4,185,787.

d. Multiplicity Adjustment to User Sampling Weight

Unlike grantees and sites, an adjustment had to be made to the user sampling weights to account for multiple chances of selection for users in the cases where such multiplicity was known and quantifiable. In addition, an adjustment was made for users that were actually selected more than once.

Prior to a multiplicity adjustment, we are dealing with the probability of selection of user *records*, not necessarily *users*. If a user is represented on a frame by more than one data record, the probability of selection for the user record does not represent the probability of selection for the user. The multiplicity adjustment to the sampling weight for user records accounts for the number of times the user could have been selected. After making such an adjustment for multiplicity, the probability of selection (and the corresponding sampling weight) will reflect users rather than user records.

The extent to which selected users could have been on the frames of more than one grantee eligible for selection (an unlikely event) is not known. Nor do we know the extent to which they could have been on the names of more than one eligible site per grantee. Across-site multiplicity should be a rare event, because the sampling instructions were designed to ensure that this did not occur among selected sites. The sampling instructions also requested that each frame be unduplicated so that each user was listed only once; however, due to the limited capabilities of some grantees' data systems, several frames may have had multiple data records for some users. Although three users from three different grantees were actually selected twice (see below), only one grantee among these acknowledged that this multiplicity *could* have occurred for other sampled cases. Only one among the other grantees, those that did not have a user selected multiple times, admitted to a multiplicity problem. We believe that there may have been multiple chances of selection for users other than those known; however, given the number of users actually selected twice, the expectation is that only a small number of users had multiple selection opportunities. In any case, we can only adjust for known multiplicity.

Because we did not obtain the sampling frames from the grantees, we could not determine for ourselves whether or not multiplicity existed, and had to rely on grantee reports. We also explored the possibility of using information collected in the questionnaire to give an indication of multiplicity. Respondents were asked how many different sites they had been to over the last three years (if the grantee had more than one site). While 209 of the 1,932 respondents said they had been to more than one site, this was not usable as an indication of multiplicity across sites, because of the three-year reference period. Respondents were also asked for all of the places in which they received medical care over the past year, and one of the possible responses was "another community, migrant, or rural clinic/center." While 61 respondents said they had been to such a place, this was not usable as an indication of multiplicity across grantees, because there was no way of knowing if this place was a 330-funded CHC.

In one grantee, all users from their school-based clinic sites were also on the list generated for the main site. Each sampled user from the two selected school-based clinics had his/her preliminary sampling weight adjusted for multiplicity. Unfortunately, it is impossible to identify any students that may have been selected from the main site's frame that were also students at one of the grantee's school-based clinics (although it is expected that the number of such students, if any, is negligible). Ideally, any such users would have had their weights adjusted similarly. The multiplicity adjustment to the user record sampling weight was as follows:

$$(14)\ W_M(U_q) = \frac{W_S(U_q)}{t}$$

where t is the number of times user U_q could have been selected. All school-based users from this grantee had two chances of selection (t=2), and one person selected twice from the same site (who also happened to be a school-based user) had three chances of selection (t=3). Each of the two users selected twice from the other two grantees had t=2. There were also four users from another grantee that had two chances of selection but were selected only once. These users had t-2. For all other users, t=1.

For the three medical users who were actually selected twice, only one record will be retained on the file for each user. The weights for these three users will be multiplied by two to account for the fact that they really represent two selections:

$$(15)\quad W_M(U_q) = 2\,W_M(U_q) = \frac{2\,W_S(U_q)}{t}$$

Note that, when t-2, this reduces to $W_S(U_q)$. *The* weight for the remaining users selected from the school-based-clinic sites (those selected only once) will be:

$$(16)\quad W_M(U_q) = W_M(U_q) = \frac{W_S(U_q)}{2}$$

And the weight for all others will be:

$$(17)\quad W_M(U_q) = W_M(U_q) = W_S(U_q)$$

At this point, the sum of the multiplicity-adjusted sampling weights is equal to 4,147,969. The minimum value for the weight is 548.3 and the maximum value is 2096.4. Fifty two cases were affected by this adjustment.

e. User-Level Analysis Weight

Several adjustments were made to the multiplicity-adjusted user sampling weights to account for nonresponse and ineligibility, as well as quantifiable undercoverage in one grantee. Because there are three distinct categories of nonresponse (users missing due to grantee nonresponse, nonlocatable users, and users located but not participating), the weighting class adjustment for nonresponse took place in three stages with different weighting classes and different assumptions about eligibility among nonrespondents. (There was no nonresponse at the site level, so a similar adjustment for site-level nonresponse was unnecessary.)

The adjustment factor for each stage is generally calculated and applied by weighting class, where the **class** is determined by factors thought to be associated with the propensity for nonresponse at that stage as well as the values of the main analytical variables of interest. An attempt was made to define the classes so that they comply with rules concerning minimum cell sizes (20 respondents per cell) and maximum adjustment factors (not greater than 2 in any cell), although these were used as guidelines rather than unbreakable rules, and were weighed against the potential bias that would result from collapsing dissimilar cells.

A poststratification adjustment was necessary in the one grantee where undercoverage was known and quantifiable. No adjustment for the undercoverage of the two excluded sites (those that opened late in 1994) was implemented because no reliable external size estimates are available, and because it is quite possible that the users of those new sites were among the counts of users from the other sites.

The preliminary analysis weights resulting from these adjustments were then examined for outliers. Trimming the weights was considered. An evaluation was carried out to determine whether the reduction in variance gained by trimming the weights would outweigh any associated bias introduced in doing so.

Adjustments to the multiplicity-adjusted sampling weight are discussed, and were implemented, in the following order:

- Weighting class adjustment for users missing due to grantee nonresponse
- Poststratification adjustment for one grantee with January through June missing from frame
- Weighting class adjustment for users missing due to screener-level nonresponse
- Weighting class adjustment for users missing due to interview-level nonresponse

Weighting Class Adjustment for Users Missing Due to Grantee Nonresponse. Two of the fifty selected grantees (and their substitutes) refused to participate: one in stratum **1** U and the other in stratum 2R. The first weighting class adjustment accounted for users lost to this grantee-level nonresponse.

Sampling strata were used as the weighting classes for this adjustment. This is an appropriate choice because the propensity for grantee nonresponse appears to be related to stratum, as evidenced by the fact that their substitutes within the same stratum also refused. For all but the two strata containing the refusers, this weighting class adjustment factor will be equal to one. For the two strata containing the refusers, the weighting class adjustment factor will be greater than one.

The weighting class adjustment factor for the estimated number of users lost to grantee-level nonresponse for stratum *h can* be thought of as the total number of users targeted in stratum *h* divided by the number of users represented by the responding grantees. This factor can be quantified as:

$$(18)\quad A_G(h) = \frac{\sum_{i=1}^{M_h} \hat{CMOS}_i}{\sum_{j=1}^{m_h} W_S(G_j)\ \hat{GMOS}_j\ \delta(G_j)} = \frac{m_h}{\sum_{j=1}^{m_h} \delta(G_j)}$$

where $\delta(G_j)$ is an indicator variable that is equal to 1 when grantee G_j participated, and equal to 0 when grantee G, refused to participate. In one stratum, containing 442 sample members, the adjustment factor was $\frac{9}{8}$. In the other stratum, containing 285 sample members, the adjustment factor was $\frac{6}{5}$.

An interim user-level weight for user U_q can now be computed as:

$$(19)\quad W_I(U_q) = W_{M'}(U_q)\, A_G(h)$$

The minimum value for this interim user-level weight is 548.3 and the maximum value is 25 15.7. The sum of these weights is 4,3 14,801.

Poststratification Adjustment for One Grantee with January - June Missing from Frame. The next step in the user weight adjustment process was poststratification. Because the totals being used to calculate these weights (grantee-specified counts of medical users at the site and grantee levels) are the best size estimates available, we did not do any sort of poststratification to *external* counts. However, for

one grantee, the frame for the subsampled site listed only those users with one or more visits during the second half of 1994. The computer system was changed on July 1, 1994, and the data from the old system were no longer available. It is believed that the users who did not show up on the frame (those who made visits *only* in the first half of 1994) are not meaningfully different than those on the frame, and so we re-weighted the users from the second half of 1994 so that they represent all users in 1994 from that site.

The 45 users from this grantee will have their weight blown up by a factor (PSADJ) of 2,485/1,508=1.65, where 2,485 is the number of users at the site as estimated by the grantee and 1,508 is the number of users on the site's July-through-December frame. The factor PSADJ will be set equal to one for users from all other grantees.

So, an interim user-level weight can be calculated as:

(20) $W_{1}(U_q) = W_{1}(U_q)\ PSADJ$

The sum of these weights is 4,368,489. The minimum value is 548.3 and the maximum value is 3028.5.

Weighting Class Adjustment for Users Missing Due to Screener-Level Nonresponse. A relatively large number of users (about 13 percent) were not locatable based on the information provided by the grantees, and despite field efforts to locate them. Sampled users lost to what we will refer to as "screener" nonresponse (although there was no screener ***per se)*** were accounted for in a weighting class adjustment. These screening nonrespondents are users for whom it was never determined whether the selected person was in fact eligible. Because the propensity for screener nonresponse and the likelihood of being found to be eligible varied greatly from one grantee to another, the grantee was used as the weighting class adjustment cell for screener-level nonresponse. (Collapsing of two grantees to form one

cell was carried out for two grantees from the same city, where one grantee had an adjustment factor greater than two.)

The weighting class adjustment factor for the number of users lost to screener nonresponse for grantee G_j can be thought of as the total number of users targeted in grantee G_j divided by the number of users represented by the sampled users with known eligibility status in grantee G_j. Let $\delta_s(U_q)$ be an indicator function that is equal to 1 if user U_q from grantee G_j had eligibility status determined (and 0 otherwise). Then the adjustment factor can be quantified as:

$$(21)\quad A_S(j) = \frac{\sum_{q=1}^{u_j} W_{I'}(U_q)}{\sum_{q=1}^{u_j} W_{I'}(U_q)\delta_s(U_q)} \quad or \quad \frac{\sum_{k=1}^{s_j}\sum_{q=1}^{u_k} W_{I'}(U_q)}{\sum_{k=1}^{s_j}\sum_{q=1}^{u_k} W_{I'}(U_q)\delta_s(U_q)}$$

where the first term is used for grantees with no site subsampling and the second for those with site subsampling. It should be noted that the "frame ineligibles" (those that were known to be ineligible prior to "screening") were given their own cell and an adjustment factor of one. Nonlocatables were given an adjustment factor of zero. All other cases (including those found to be ineligible as part of screening) were assigned the adjustment factor above.

A new interim user-level weight can now be computed:

$$(22)\quad W_{I''}(U_q) = W_{I'}(U_q)\ A_S(j)$$

Among those with nonzero weights, this weight had a minimum value of 780.9 and a maximum value of 3223.5. (There were 392 cases with zero weights.) The sum of these weights equals 4,368,489.

Weighting Class Adjustment for Users Missing Due to Interview-Level Nonresponse. Because the propensity for interview nonresponse and the values of the outcomes of interest are likely to be affected by age, sex, and geography, we used stratum by sex by age group (0- 17, 18-40, 41+) by urban/rural by

Census Region[6] as the initial weighting class adjustment cell for interview nonresponse. Note that urban/rural crossed with Census Region is the same as the sampling strata, except that the managed care stratum grantees and the certainty stratum grantee were put back with their geographically-similar counterparts. Collapsing across Census Region became necessary only once: for the cells corresponding to males ages 18-40 in the Northeast and Midwest Census Regions, where the total number of respondents in each was less than ten After this collapsing, there were 53 cells: 3 7 cells had 20 or more respondents; 12 cells had 15-19 respondents; and four cells had 11-14 respondents.

The weighting class adjustment factor for the number of users lost to interview nonresponse for class c can be thought of as the total number of users targeted in class c divided by the number of users represented by the responders[7] in class c. It can be quantified as:

$$(23)\quad A_I(c) = \frac{\sum_{q \in SE_c} W_{I'}(U_q)}{\sum_{q \in SE_c} W_{I'}(U_q)\delta_I(U_q)}$$

where:

$\delta_I(U_q)$ is an indicator variable that is equal to 1 if user U_q responded to the interview, and 0 otherwise,

SE_c is the set of all sampled eligible users in class c.

For this adjustment, nonlocatables and ineligibles (both frame- and screener-determined) were assigned to their own cells and given a factor of one. Eligible noncompletes were assigned a factor equal to zero. Eligible completes were assigned the above factor.

---

[6]The two rural South Census Region groups that were divided into two strata for sampling purposes, based on Census Division, were kept separate when creating these cells.

[7]Responders include the 1,924 completes plus eight partial completes that were judged to be completes.

In combining the three weighting class adjustment factors for nonresponse with the multiplicity-adjusted user sampling weight, the final analysis weight (aside from any trimming) for a user U_q from grantee G_j and also in class c is:

(24) $W_A(U_q) = W_{1'}(U_q)\ A_I(c)$

Among those with nonzero weights, this weight had a minimum value of 805.2 and a maximum value of 4028.4. (There were 63 1 cases--nonlocatables and nonrespondents--with zero weights.) The sum of these weights is equal to 4,368,488.5. Among the 1,932 completes, the weights ranged from 837.8 to 4028.4, with a sum of 3,802,768.

f. Truncation

We examined outlier weights defined three ways: (1) below the 1st percentile and above the 99th percentile; (2) below the 5th percentile and above the 95th percentile; and (3) below the 10th percentile and above the 90th percentile. Legitimate reasons were found for the upper and lower tails of the weight distribution. To determine whether trimming the weights would be beneficial, we examined 83 critical variables from the user survey, about half of which are conditions from the Condition List section of the questionnaire. We calculated a mean square error as follows:

(25) $MSE = \textbf{\textit{variance}} + bias^2$

where:

variance is the square of the standard error of the estimate as output by SUDAAN (see next section), and ***bias*** is defined as the difference between the estimate using the untruncated weight and the estimate using the truncated and smoothed weight. The goal is to minimize the mean square error, which will rise with an increase in variance (due to unequal weighting) or an increase in bias (due to truncation). Taking the average mean square error over all 83 estimates of critical variables, we found:

- average mean square error for untruncated weights is 1.84403 1
- average mean square error for weights truncated at 1 percent and 99 percent is 1.845 186
- average mean square error for weights truncated at 5 percent and 95 percent is 1.847858
- average mean square error for weights truncated at 10 percent and 90 percent is 1.879718

Because the truncation increased the mean square error in all three instances, a decision was made not to use truncated weights.

g. Variance and Design Effects

The variance of estimates from the user survey were calculated using SUDAAN software, which is specifically designed to analyze survey data resulting from samples with complex designs. After consulting with the software's primary author, we used the with-replacement sample design option, which assumes:

- with-replacement sampling at the first stage (or sampling fractions of less than ten percent in every first stage stratum)
- sampling with or without replacement at subsequent stages, and
- sampling with equal or unequal probabilities of selection at the first and subsequent stages.

For such a design, SUDAAN uses the between-PSU (grantee) within-stratum variance component to estimate the variances. This is the option recommended for the more complicated sample designs. Because it does not take into account the between-site within-grantee variance, this option results in a conservative (high) estimate of the variance. Table II. 14 presents estimates and standard errors for the critical variables.

The design effect of an estimate measures the impact of a complex sample design on the variance of the estimate. It is calculated as the variance estimate accounting for design complexities divided by the variance one would have achieved using a simple random sample of the same size. The component of the

TABLE II. 14

ESTIMATES AND STANDARD ERRORS FOR USER SURVEY
(Critical Variables)

	Variable Number and Name	Number Responding	Unweighted Frequencies	Weighted Estimate (percent)	Standard Error (SUDAAN)
c1a	Stiffness foot, leg,...,back	1930	**131**	6.89	**0.86**
c1b	Arthritis	1918	**347**	**18.13**	**1.49**
c1c	Neck, back, spine	1931	**284**	**14.59**	**1.62**
c1d	Dermatitis	1932	**149**	**7.46**	1.01
c2a	Deafness	1925	**56**	**2.99**	0.43
c2b	Trouble hearing	1927	**124**	**6.61**	0.69
c2c	Blindness	1931	**37**	**1.97**	0.30
c2d	Trouble seeing	1925	**136**	**6.96**	0.97
c2e	(Cerebral) palsy	1924	**5**	**0.25**	0.13
c2g	Accident/injury	1926	**147**	**7.42**	1.14
c3a	Gallbladder	1171	**30**	**2.55**	0.53
c3b	Cirrhosis	1172	**8**	**0.79**	0.29
c3c	Hepatitis	1926	**13**	**0.77**	0.22
c3d	Ulcer	**1170**	**70**	**5.87**	0.67
c3e	Indigestion	**1930**	**176**	**9.13**	0.81
c3f	Other stomach	**1930**	**133**	**6.62**	0.90
c3g	Other digestive	**1929**	**44**	**2.11**	0.44
c4a	Goiter/thyroid	**1925**	**36**	**1.81**	0.28
c4b	Diabetes	**1924**	**116**	**6.23**	0.67
c4c	Anemia	**1916**	**116**	**5.82**	0.66
c4d	Epilepsy/seizures	**1931**	**35**	**1.86**	0.37
c4e	Kidney	**1927**	**80**	**4.04**	0.53
c4f	Bladder	**1927**	**116**	**5.78**	0.70
c4g	Genital	**1929**	**32**	**1.63**	0.28
c4h	Prostate	**335**	**10**	**3.40**	1.39
c4i	Breast cancer	**839**	**3**	**0.42**	0.24
c4j	Uterus/ovaries	**833**	**30**	**3.28**	0.67
c4k	Other female	**1254**	**83**	**6.63**	1.02
c5a	Arteriosclerosis	**1171**	**18**	**1.77**	0.41
c5b	Congenital heart disease	**1918**	**32**	**1.61**	0.30
c5c	Coronary heart disease	**1163**	**25**	**2.29**	0.80

TABLE II. 14 (continued)

Variable Number and Name		Number Responding	Unweighted Frequencies	Weighted Estimate (percent)	Standard Error (SUDAAN)
c5d	Hypertension	1921	395	21.17	1.47
c5e	Stroke	1929	33	1.73	0.32
c5f	Angina	1168	40	3.63	0.61
c5g	Heart attack	1173	48	4.24	0.79
c5h	Cancer	1930	59	3.27	0.40
c5i	Lead poisoning	605	10	1.76	0.59
c5j	Other heart trouble	1929	54	2.99	0.43
c5k	Other circulatory	1921	64	3.58	0.49
c6a	Bronchitis	1924	217	11.11	1.11
c6b	Asthma	1925	183	9.53	0.90
c6c	Emphysema	1171	26	2.41	0.50
c6d	Tuberculosis	1927	10	0.51	0.16
c6e	Tonsillitis	1929	111	5.50	0.56
c6f	Work-related respiratory	1144	6	0.52	0.21
c6g	Other respiratory	1928	72	3.77	0.44
e3a	Hispanic ethnicity	1926	562	26.86	4.00
e4a & e4b	Race-white	1852		49.26	4.01
	Race-black			33.35	3.93
	Race-native American			1.90	0.66
	Race-Asian/Pacific			3.66	2.33
	Race-Other			11.83	3.01
e5a	Working	1173	498	41.45	1.84
e8	Marital-married, sps in hh	1321	480	35.83	1.72
	Marital-married, sps away		21	1.66	0.45
	Marital-widowed		114	9.13	1.06
	Marital-divorced		166	12.90	1.22
	Marital-separated		74	5.62	0.67
	Marital-never married		466	34.85	1.75
e10a	Country of birth-P.Rico	350	55	17.65	4.46
	Country of birth-Virgin Isl.		2	0.46	0.35
	Country of birth-Guam		0	0.00	0.00
	Country of birth-Canada		5	1.28	0.91
	Country of birth-Cuba		2	0.64	0.43
	Country of birth-Mexico		184	47.70	5.55
	Country of birth-other		102	32.27	6.99

TABLE II. 14 (continued)

	Variable Number and Name	Number Responding	Unweighted Frequencies	Weighted Estimate (percent)	Standard Error (SUDAAN)
m0	CHC usual source	1926	1605	83.94	1.46
m1a	If not, have usual source	322	281	87.50	1.94
	No usual source		30	8.69	1.45
	More than one usual srce		11	3.81	1.34
m1b	Place goes to most often	11	6	52.27	17.44
m5ba	Center convenient	1599	728	45.56	4.07
m5bb	Hours convenient		264	17.12	2.99
m5bc	Don't have to wait		144	9.32	1.76
m5bd	Know and trust		502	30.69	3.19
m5be	Kind of care		314	19.94	2.89
m5bf	Afford		292	17.37	2.61
m5bg	Medicaid		157	9.95	1.56
m5bh	Your insurance		114	7.32	1.48
m5bi	Child care		7	0.48	0.19
m5bj	Transportation		21	1.37	0.40
m5bk	Language		67	3.85	1.35
m5bl	Care		260	15.99	2.28
m5bm	No other place		120	7.42	1.55
p1	Hypertension-yes	1556	381	25.14	1.75
	Hypertension-no		1147	73.06	1.77
	Hypertension-pregnancy		13	0.81	0.25
	Hypertension-borderline		15	0.99	0.34
p10	Cholesterol	916	264	28.20	2.30
p12	Diabetes	1556	116	7.69	0.80
p25a	Asthma	1815	174	9.64	0.91
u1a	Bathing	1172	40	3.79	0.71
u1b	Dressing	1174	42	3.76	0.75
u1c	Eating	1174	15	1.30	0.46
u1d	Transfer	1174	39	3.54	0.73
u1e	Toileting	1174	22	1.99	0.51
u1f	Getting around house	1174	26	2.23	0.59
x1	Medicare	1915	234	12.76	1.34
x2	Medicaid	1903	731	39.08	2.53
x3	Other public insurance	1905	101	5.37	1.28
x5	Military insurance	1916	29	1.72	0.43
x6	Private insurance	1898	509	27.03	2.49

TABLE Il. 14 *(continued)*

Variable Number	Name	Number Responding	Unweighted Frequencies	Weighted Estimate (percent)	Standard Error (SUDAAN)
z1a	Ever had pap	826	785	95.06	0.83
z3a	Heard of mammogram	555	489	88.48	1.77
z3b	Ever had mammogram	489	359	73.43	2.39
				(mean)	
e2a	Education	1652		9.07	0.18
e9b	Family income	1158		10.07	0.19
z1bmon	Most recent pap	745		16.77	1.67
z3dyr	Most recent mammogram	341		1.85	0.20

design effect due to unequal weighting was only 1.08. Achieving such a small unequal weighting component is the result of our efforts during the sampling process to equalize probabilities of selection across all grantees, to the extent possible.

Table II. 15 shows the design effects for the critical variables. The estimated design effects for the Condition List variables ranged from 0.85 (c4a: goiter/thyroid) to 4.07 (clc: neck/back/spine), with an average of 1.66. This means that the complex design increased the variance of estimates in this section by two-thirds, resulting in an effective sample size of 1 ,164 (compared to the nominal sample size of 1,932). For the other key variables, there were some large design effects for variables related to the clustering in the design; that is, related to characteristics of grantees. These included race (design effects range from 4.29 [Native American] to 28.43 [Asian]), whether Hispanic (15.69) country of birth (particularly Puerto Rico [4.65], Mexico [4.19], and other [7.57]), center characteristics (question series m5, particularly if the center is convenient [10.75], the hours are convenient [10.13], and kind of care [8.43]), first type of place mentioned where received care (grantee [14.98], other clinic [6.83]), and some types of insurance status (Medicaid [5.11], other public insurance [6.18], and private insurance [5.98]). Demographic characteristics, such as employment status, marital status, family income, chronic diseases, limitations in activities of daily living, and cancer screening all had design effects of less than two.

While these design effects may be slightly overstated, due to the conservative approach used to calculate the variances, selecting a larger number of grantees would mitigate the larger design effects for grantee-related variables, such as race and Hispanic ethnicity.

These design effect estimates can be compared to those estimated by MPR when the sample design was being developed. Using various assumptions for between-grantee and between-site variance proportions, and assuming a three-stage design with 2,000 respondents and 70 sites, we came up with design effect estimates of 7.027 (assuming 3.3 percent of the total variance was accounted for by the

TABLE II. 15

DESIGN EFFECTS FOR UNTRUNCATED WEIGHTS
(Critical Variables)

	Variable Number and Name	Design Effect: By Estimate	Design Effect: By Question or Section	Effective Sample Size (nominal n=1,932)
c1a	Stiffness foot, leg,...,back	2.22	1.66	1162.79
c1b	Arthritis	2.88		
c1c	Neck, back, spine	4.07		
c1d	Dermatitis	2.85		
c2a	Deafness	1.20		
c2b	Trouble hearing	1.50		
c2c	Blindness	0.91		
c2d	Trouble seeing	2.81		
c2e	(Cerebral) palsy	1.20		
c2g	Accident/injury	3.63		
c3a	Gallbladder	1.34		
c3b	Cirrhosis	1.27		
c3c	Hepatitis	1.17		
c3d	Ulcer	0.97		
c3e	Indigestion	1.51		
c3f	Other stomach	2.52		
c3g	Other digestive	1.81		
c4a	Goiter/thyroid	0.85		
c4b	Diabetes	1.48		
c4c	Anemia	1.53		
c4d	Epilepsy/seizures	1.45		
c4e	Kidney	1.40		
c4f	Bladder	1.74		
c4g	Genital	0.97		
c4h	Prostate	2.04		
c4i	Breast cancer	1.13		
c4j	Uterus/ovaries	1.17		
c4k	Other female	2.10		
c5a	Arteriosclerosis	1.16		
c5b	Congenital heart disease	1.13		
c5c	Coronary heart disease	3.34		
c5d	Hypertension	2.49		

TABLE II. 15 *(continued)*

	Variable Number and Name	Design Effect: By Estimate	Design Effect: By Question or Section	Effective Sample Size (nominal n=1,932)
c5e	Stroke	1.19		
c5f	Angina	1.27		
c5g	Heart attack	1.82		
c5h	Cancer	0.97		
c5i	Lead poisoning	1.22		
c5j	Other heart trouble	1.23		
c5k	Other circulatory	1.34		
c6a	Bronchitis	2.38		
c6b	Asthma	1.81		
c6c	Emphysema	1.26		
c6d	Tuberculosis	0.96		
c6e	Tonsillitis	1.15		
c6f	Work-related respiratory	0.96		
c6g	Other respiratory	1.03		
e3a	Hispanic ethnicity	15.69	15.69	123.14
e4a & e4b	Race-white	11.91	14.71	131.37
	Race-black	12.87		
	Race-native American	4.29		
	Race-Asian/Pacific	28.43		
	Race-Other	16.03		
e5a	Working	1.65	1.65	1170.91
e8	Marital-married, sps in hh	1.70	1.63	1185.28
	Marital-married, sps away	1.65		
	Marital-widowed	1.78		
	Marital-divorced	1.75		
	Marital-separated	1.11		
	Marital-never married	1.79		
e10a	Country of birth-P. Rico	4.65	3.42	565.46
	Country of birth-Virgin Isl.	0.90		
	Country of birth-Guam		--	
	Country of birth-Canada	2.21		
	Country of birth-Cuba	0.98		
	Country of birth-Mexico	4.19		
	Country of birth-other	7.57		

TABLE Il. 15 *(continued)*

Variable Number and Name		Design Effect: By Estimate	Design Effect: By Question or Section	Effective Sample Size (nominal n=1,932)
m0	CHC usual source	3.03	1.58	1225.89
m1a	If not, have usual source	1.07		
	No usual source	0.82		
	More than one usual srce	1.52		
m1b	Place goes to most often	1.44		
m5ba	Center convenient	10.75	6.39	302.17
m5bb	Hours convenient	10.13		
m5bc	Don't have to wait	5.92		
m5bd	Know and trust	7.72		
m5be	Kind of care	8.43		
m5bf	Afford	7.64		
m5bg	Medicaid	4.38		
m5bh	Your insurance	5.23		
m5bi	Child care	1.22		
m5bj	Transportation	1.95		
m5bk	Language	7.89		
m5bl	Care	6.26		
m5bm	No other place	5.60		
P1	Hypertension-yes	2.53	2.02	958.81
	Hypertension-no	2.47		
	Hypertension-pregnancy	1.19		
	Hypertension-borderline	1.87		
p10	Cholesterol	2.47	2.47	782.19
P12	Diabetes	1.41	1.41	1370.21
p25a	Asthma	1.72	1.72	1123.26
u1a	Bathing	1.63	1.79	1081.34
u1b	Dressing	1.85		
u1c	Eating	1.98		
u1d	Transfer	1.83		
u1e	Toileting	1.56		
u1f	Getting around house	1.87		
x1	Medicare	3.10	4.50	429.14
x2	Medicaid	5.11		
x3	Other public insurance	6.18		
x5	Military insurance	2.14		

TABLE II. 15 (continued)

Variable Number and Name		Design Effect: By Estimate	Design Effect: By Question or Section	Effective Sample Size (nominal n=1,932)
x6	Private insurance	5.98		
z1a	Ever had pap	1.23	1.23	1570.73
z3a	Heard of mammogram	1.73	1.60	1207.50
z3b	Ever had mammogram	1.47		
e2a	Education	2.62	2.62	737.40
e9b	Family income	1.61	1.61	1200.00
z1bmon	Most recent pap	1.63	1.63	1185.28
z3dyr	Most recent mammogram	1.51	1.51	1279.47

between-grantee component and 29.7 percent by the between-site within-grantee component) and 3.270 (assuming 3.3 percent between-grantee and 6.7 percent between-site within-grantee).

D. VISIT SURVEY

1. Sample Implementation

a. Process and procedures

All grantees were sent instructions (see Appendix 6) on how to create the sample frame of medical encounters. The first step was to read the instructions carefully and then call their assigned field coordinator at MPR. The grantee was instructed to compile one list of all medical encounters at any of their eligible sites in calendar year 1994. To the extent possible, they were asked to remove any known ineligible encounters, such as non-medical encounters and nurse-only encounters, and to remove any duplicate entries of medical encounters. The lists of eligible medical encounters were generally in machine-readable format, but some grantees' systems could generate usable lists only on paper. One grantee compiled its list manually by looking at daily visit logs. When the list was compiled, the grantee was instructed to call the field coordinator once again to give the total number of records on the list.

At this point, the field coordinator gave the frame size to the Sampling Manager, who then systematically selected the encounter list numbers for the sample. These numbers were sent via fax to the grantee. Previously, either when the list was first generated or after the frame size was given to MPR, the grantee was asked to sort (re-order) the list chronologically, if possible. The grantees were also instructed to number their encounter lists consecutively, beginning with the number one. When the grantees received the selected encounter list numbers, they were asked to provide as much of the following information as was available on each of the encounters corresponding to the selected list numbers: list number, date of encounter, medical record number, usual location of medical record, patient's date of birth, patient's sex, site of encounter, and something to uniquely identify the encounter if the patient had more than one medical encounter that day.

As described more fully in the Sample Design section, encounter sampling took place in two stages. The initial sample was supplemented by additional sample later in the field period, owing to yields that were lower than expected. When the additional sample was selected by the Sampling Manager, the selected numbers were sent by fax to the grantee, who was then instructed to go back to the same list of eligible medical encounters and provide the locating and other information on each of the encounters corresponding to the selected list numbers.

Once the locating and other information was received for the sample, the cases were entered into a database, assigned study identification numbers, and set up on the receipt and tracking system. Abstraction form labels were then generated.

In this manner, a total of 3,195 cases were selected: 2,763 in the initial sample and 432 in the additional sample. The grantees were offered an incentive payment of $100 if they prepared the encounter sample frame by June 26, 1995 (and $50 to $75 if by July 6, 1995). Fourteen of the grantees did not create the encounter sample frames themselves, but used the services of their data processing vendor. Nine of these 14 grantees had service provided by a single vendor who did not charge for the service. The remaining five vendors charged a fee ranging from $150 to $533, for a combined total of $1,853. Note that an additional payment of $250 was made to those grantees that completed the medical encounter chart abstractions using their own staff.

b. Timeframe

Encounter samples were selected on a flow basis immediately after each frame size was received from the grantee. The first encounter sample was selected on June 14, 1995. About one-third of the initial samples were selected in June 1995, and about half were selected in July, August, and September of 1995. Four grantees had their initial samples selected in October, two in November, and one in early February of 1996. The additional encounter samples were selected on September 19 and 20, 1995.

Obtaining the encounter frame counts from the grantees took longer than it did for the user samples for most of the grantees, even with the incentive offered for early completion of the sample frame. As with the user survey, some had staff shortages and others had systems limitations (which were exacerbated by the fact that these frames were generally three times the size of the user frames). Some could not maintain such a large file on their systems and had to print out the file on paper. Staff turnover at all levels of the CHCs, and even at one of the data processing vendors, was a large problem for the encounter component, because more time had elapsed since the grantee had initially agreed to participate in the study. When a delay occurred in the sampling process, it was often due to the departure of a staff member who had been responsible for assisting either directly or indirectly with sample selection. Some grantees, and even some vendors, had to generate the frame more than once because the first one had been generated incorrectly. Obtaining the frames from most of the vendors took an inordinate amount of time. Once the frames were generated and the samples selected there was a long delay for some grantees to provide the corresponding chart locating information for the selected cases.

Two of the grantees who had additional sample selected could not provide locating information for their selected cases because they could not locate the original encounter frame and could not generate another easily. Because we wanted to close down the field period, it was decided in mid-December 1995 that these twenty additional cases should not be released and hence should not be considered part of the selected sample. That left 3,175 medical encounters selected in the sample.

c. Problems

In addition to the problems mentioned above and in the Sample Design section, there were others worth noting. Many grantees could not sort their frames chronologically, many could not provide one list covering all eligible sites, and some could not consecutively number their lists. When unable to number, most did find a way to find the selected cases, such as counting the number of cases per page or screen, locating the page or screen on which the selected case could be found, and then counting out which record

on the page or screen was the selected one. For one grantee, however, this proved to be an extremely difficult problem to overcome. The encounters were supposedly numbered consecutively as they occurred (not starting with the number one), although the counter reverted to 00001 after it reached 99999. The first problem was that, after the removal of known ineligible encounters, the numbers were no longer consecutive. After ineligible encounters were removed, the number of records was known; therefore, the plan was to oversample accordingly, assuming a certain "hit" and "miss" rate for selected encounters. The second problem turned out to be that the encounter numbers had not been assigned consecutively after all, which rendered the encounter numbers useless for our purposes. We reverted to a tedious but productive method whereby one of our interviewers sat at a computer screen at the CHC and scrolled through the list of eligible encounters, counting out the encounters until he reached the selected cases.

Most grantees could not remove all types of ineligible encounters from their frames, but many could remove most types of known ineligible encounters (or did not have such encounters on their systems in the first place). Several grantees had problems with the concept of an encounter-level list. We received a few frames that turned out to be user-level lists, one of them having encounter counts on the user record, which had to be re-run. Some, including at least one vendor, did not comprehend the notion of a list of encounter records, and thought more in terms of reports with aggregate counts. This confusion caused the actual number of selected encounters to be lower than what was targeted for many grantees, even after the additional sample cases had been selected. Built into the initial and additional sample size formulas was the anticipated number of encounters on the sample frames, based on estimates given to us by the grantees themselves. In a couple of instances, the grantee's estimate was much higher than the number ultimately on the frame, presumably because of confusion on the part of the grantee about what was meant by counting all medical encounters in 1994.

As mentioned above, many grantees could not provide a single list comprising encounters from all eligible sites for the entire year. Most of these provided a separate list for each site. Because of space

limitations, one provided the frame by month. One was able to provide only a manual list of encounter counts by day within site. We were able to deal with these by converting the selected list numbers to the individual list numbers, each of which started with the number one.

As with the user survey, other types of problems concerned grantees' having antiquated and/or inflexible computer systems. Several grantees upgraded their computer systems mid-year, usually completely changing systems (and data processing vendors in some cases). When this happened, most grantees tried to piece together a frame from the entire year, but two grantees could provide a frame of medical users only for part of 1994: one only from July through December of 1994, and one only from May through December of 1994, omitting any medical encounters that took place in the first part of the year. As with the user survey, some were put in the awkward position having to deal with data processing vendors that no longer provided service to them, but who had provided service in 1994.

Some grantees listed "medical encounters" that turned out not to be medical encounters at all. In a few cases, the medical encounter corresponded to one in which the patient never showed up or left before s/he was actually seen by the medical provider. The encounter dates often did not exactly match the date of the encounter as found in the medical chart; at some grantees, the date found on the system corresponded to the date the encounter form was generated (sometimes the day before the encounter) or even the date the encounter information was entered into the system (which could have been months after the encounter took place).

Of the 3,175 selected medical encounters, 23 5 cases were found to be ineligible after an attempt was made to locate the information, and 62 cases were nonlocatable or unavailable and hence had unknown eligibility status.

No encounters were selected more than once in our sample, and no grantees admitted to having had a problem with duplicate listings in their encounter frames. One frame, provided to us by a vendor, did come with duplicates, but we were able to remove the duplicates before the sample was selected.

d. Recommendations

As stated above for the user survey, for any similar endeavors carried out in the future, these problems are less likely to arise because grantees are upgrading their computer systems and thus their capabilities. The recommendation described above under the user survey applies here as well: have the grantees supply the survey contractor with the sample frames, and have the frame examined, unduplicated, and sorted by the contractor prior to sample selection. A second recommendation for the encounter sample would be to link it to the user survey sample, which would cut the sampling effort in half (for both the grantees and the contractor), allow for greater analytical possibilities, validate the user survey results, and support an analysis of potential nonresponse bias, while providing comparable effective sample sizes to those yielded by the current design.

2. Fielding of the Visit Survey

Based on preliminary information HRSA collected in the early planning stages for this survey, it was believed that because of confidentiality concerns related to the medical records, grantees would prefer to have their own staffs, rather than MPR's, abstract information from medical records, despite MPR's assurances of confidentiality. However, once the process began, grantee concerns about staff availability and burden predominated. Originally, it was thought that at least 80 percent of the grantees would prefer to have their own staffs conduct the abstractions. As it turned out, about half the encounter abstraction forms were completed by grantee staff and half by MPR staff

In general, grantee staff were better able than the MPR interviewers to complete the abstraction from records. This is evidenced by the higher degree of missing information on the MPR-completed forms when they originally reached and were reviewed by the home office. Grantee staff were more familiar with the records, terminology, and with the handwriting and names of providers. For example, the provider type often had to be determined by the name signed in the record. Staff were much more familiar with this type

of information than interviewers, and assisted the interviewers when necessary. Efforts were made to collect as much of the originally missing information as possible.

The fielding of the records abstractions extended from May 1995 through March 1996, though the original collection was planned to take only four months. The reasons for the delays are the same as those for the user survey: technical problems, staff shortages, staff turnover, and lack of time on the part of the grantees' staffs (some had to do the abstraction on weekends).

3. Response Patterns

a. Response rate

The overall completion rate among eligibles for the encounter abstraction component of the study was 98.0 percent. The final status for all selected cases can be found in Table U.16. Of the 3,175 medical encounter cases selected and released, 2,878 were known to be eligible, and all of these were completed. There were no refusals. Of the 2,878 completes, 1,491 (52 percent) were completed by a staff member at the community health center and 1,3 87 (48 percent) were completed by an MPR interviewer.

The remaining cases were either ineligible (n=235) or cases where the patient's medical chart could not be located or was unavailable during the field period and, therefore, the eligibility of these cases could not be confirmed (n=62). One of these 62 cases was partially completed from information on electronic records, but the chart could not be located to complete the abstraction form. The 23 5 ineligible cases were ones where the selected record did not correspond to a medical encounter or did not meet the study's eligibility criteria for medical encounters. Forty-five of these 235 ineligible cases were associated with abstraction forms that had been completed but were later found to be ineligible, primarily because neither a physician nor a mid-level practitioner had been seen (n=34), or because the visit did not take place in 1994 (n=5).

Completion rates did not vary by urbanicity, and varied little by region. Those selected from rural CHCs were just as likely to be completed as those from urban CHCs (98.0 percent versus 98.1 percent).

TABLE II. 16

ENCOUNTER ABSTRACTION STUDY FINAL STATUS

	Number of Cases	Percent of Selected Sample	Percent of Completes
Total sample selected	3,175	100.0%	--
Ineligible:	235	7.4	—
Encounter did not meet eligibility criteria	231	7.3	--
Record did not correspond to an encounter	4	0.1	--
Nonlocatable:	62	2.0	--
Record not available at close	7	0.2	--
Record not located at close	55	1.7	--
Completes:	2,878	90.6	100.0%
By CHC staff	1,491	47.0	51.8
By MPR interviewer	1,387	43.7	48.2

Encounters selected from CHCs located in the Northeast were most likely to be completed (98.7 percent), while those selected from a Midwestern CHC were the likely to respond (97.3 percent), due to the larger percentage of ineligible visits sampled.

The overall response rate is calculated here as the number of eligible completes divided by the estimated number of eligible encounters. The estimated number of eligible users is the number of known eligibles plus a fraction of those with unknown eligibility status. The fraction used is the eligibility rate as determined from those with known eligibility status. The eligibility rate is estimated at ,925; that is, the number of known eligibles (2,878) divided by known eligibles plus known ineligibles (2,878 + 235). This estimate assumes that the encounters with nonlocatable charts have the same eligibility distribution as those that were located. Applying this rate to the nonlocatables yields an estimate of 57 eligibles out of 62. The overall response rate can then be calculated as 2,878/(2,878+57) = ,980, or 98 percent.

b. Item nonresponse and other concerns

Table II. 17 shows the number of missing responses for all items on the abstract form, overall and by type of abstractor. The items with the highest proportion of missing responses' are race (16.8 percent missing), ethnicity (33.8 percent), and whether the patient was referred by another provider (6.8 percent). All other items had nonresponse rates of four percent or less. Because, in general, there was only one type of abstractor per grantee, any differences in patterns by type of abstraction (CHC staff member vs. MPR abstractor) may be due to differences in grantee record-keeping practices rather than the type of abstractor. For grantees in which the abstractor was a CHC staff member, the item nonresponse rates were lower for race, ethnicity, day of visit, and reimbursement type. Encounters at grantees in which MPR abstractors

[8]While questions 2 1 a, 21 b, and 23 have high *proportions* of nonresponse, there were only 11 cases for which 2 1 a and 23 should have been answered (with one nonresponding case), and only 9 cases for which 21b should have been answered (and two nonresponding cases).

TABLE II. 17

ITEM NONRESPONSE BY TYPE OF ABSTRACTOR

	Overall	Abstracted by CHC Staff Member	Abstracted by MPR Interviewer
Date of Visit	1 year (0.0%) 2 month (0.1%) 90 day (3.1%)	1 year (0.0%) 1 month (0.1%) 2 day (0.1%)	0 year (0.0%) 1 month (0.1%) 88 day (6.3%)
Date of Birth (Q2)	0 year (0.0%) 13 month (0.5%) 13 day (0.5%)	0 year (0.0%) 11 month (0.7%) 11 day (0.7%)	0 year (0.0%) 2 month (0.1%) 2 day (0.1%)
Sex (Q3)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Race (Q4)	484 (16.8%)	199 (13.3%)	285 (20.5%)
Ethnicity (Q5)	974 (33.8%)	393 (26.4%)	581 (41.9%)
Reimbursement (Q6)	83 (2.9%)	12 (0.8%)	71 (5.1%)
Source of Payment (Q7)	17 (0.6%)	9 (0.6%)	8 (0.6%)
Referred (QS)	195 (6.8%)	101 (6.8%)	94 (6.8%)
Seen Before (Q 11 A)	116 (4.0%)	73 (4.9%)	43 (3.1%)
For this Condition (Q 11B)	92 (3.8% of Q11A=1)	36 (2.9% of Q11A=1)	56 (4.7% of Q11A=1)
Disposition (Q 13)	18 (0.6%)	12 (0.8%)	6 (0.4%)
Providers Seen (Q 14)	19 (0.7%)	7 (0.5%)	12 (0.9%)
Diagnostic/Other Procedures (Q15)	22 (0.8%)	16 (1.1%)	6 (0.4%)
Counseling/Other Services (Q 16)	52 (1.8%)	36 (2.4%)	16 (1.2%)
Another Encounter (Q 17)	84 (2.9%)	69 (4.6%)	15 (1.1%)
2nd: Referred (Q 18)	0 (0.0% of Q17=1)	0 (0.0%)	0 (0.0%)
2nd: Seen Before (Q2 1A)	1 (9.1% of Q17=1)	0 (0.0%)	1 (16.7%)
2nd: For this Condition (Q21 B)	2 (22.2% of Q17=1 and Q2 1A=1)	1 (25.0% of Q21A=1)	1 (20.0% of Q21A=1)
2nd: Disposition (Q23)	1 (9.1% of Q17=1)	1 (20.0%)	0 (0.0%)
2nd: Providers Seen (Q24)	0 (0.0% of Q17=1)	0 (0.0%)	0 (0.0%)

were used had higher item response levels for the month and day of birth, the types of procedures and services received, and whether the patient had a second medical encounter that day.

There was only one question (1 lb) pertaining to the sampled encounter that had a potential logical skip, meaning that the question was intentionally not asked because it does not apply. In addition, the entire third page of the abstraction form (for the cases in which a second medical encounter occurred on the same day) is left blank if there was only one encounter. (A logical skip comparable to 1 lb, 21 b, is also found on the third page.) Logical skips are coded as -2. When calculating estimates or constructing new variables, it will generally be necessary to convert the "don't know"(-1) and "missing data" (-5) responses to a missing value (in SAS, that value 'would be a dot for numeric variables) and perhaps statistically impute a non-missing value to such cases using a procedure such as hot deck imputation. For descriptive purposes, it may be desirable to leave the "don't know" and "refused" categories intact.

There are other types of item nonresponse for coded and constructed variables. When the coder was unable to assign an ICD9 code to a particular condition because of insufficient information, a code of 7998 or 7999 was assigned. When she was unable to assign a patient complaint code from what was reported on the abstract form, a code of 89900, 89980, or 89990 was assigned. Similarly, when the coder was unable to assign a medication code, a code of 99980 or 99999 was assigned.

4. Weighting Methodology and Design Effects

This section outlines the strategy for creating sampling and analysis weights for the encounter abstraction data As described above in the section on weighting the user survey data, a *sampling weight* is the reciprocal of the probability of selection for each sampled unit. An *analysis weight* is a sampling weight that has been adjusted for nonresponse, and possibly poststratified to known population totals and/or trimmed to mitigate the effect of outlier weights.

Because both the user and visit components took place within the same set of selected clusters and grantees, cluster and grantee weight components were discussed above in the section on weighting the user

survey (II.C.4). After the grantee stage, the designs of the two study components diverge. No site subsampling took place for the Visit Survey. The third and final step in the Visit Survey sampling process was the selection of medical encounters within grantee. Medical encounter records were selected with equal probability within grantee.

Replicate samples of encounters were selected at the same time as the original samples, in anticipation of the possibility of lower yields than expected. Random selections from these replicates were released as needed during the field period to supplement the sample. For two grantees, these random selections from the replicates were never released for data collection due to time constraints. Below is a discussion of the selection of encounters and weighting class adjustments for encounters.

a. Encounter-Level Sampling Weight

The conditional probability of selection for encounter record V_r given that grantee G_j is selected can be quantified as:

$$(26)\quad P(V_r|G_j) = \frac{v_j}{GNOV_j}$$

where:

v_j is the number of encounter records selected and released for grantee G_j

$GNOV_j$ is the actual number of encounter records on the grantee G_j frame.

Note that v, includes both the original and replicate samples selected and excludes the 20 cases never released in two grantees due to time constraints.

Note that v_j and $GNOV_j$ contain encounter records known to be ineligible prior to attempts to locate the medical record as well as records later found to be ineligible. For one grantee, $GNOV_j$ represents the range of numbers from which the sample numbers were selected, not the number of records on the frame.

This was done to accommodate the grantee's inability to re-generate consecutive record numbers after removing ineligibles from the frame.

The unconditional probability of selection for encounter record V_r within grantee G_j is:

$$(27)\quad P(V_r) = P(G_j)\,P(V_r|G_j)$$

where $P(G)$, the probability of selection of grantee G_j, has been defined previously as:

$$(28)\quad P(G_j) = P(C_i)\,P(G_j|C_i) = \frac{m_h\, G\hat{M}OS_j}{\sum_{z=1}^{M_h} C\hat{M}OS_z}$$

Thus, the sampling weight for encounter record V_r selected from grantee G_j is:

$$(29)\quad W_S(V_r) = \frac{1}{P(V_r)}$$

At this point, the sampling weight has a minimum value of 2,488.20 and a maximum value of 5,870.82. The sum of these sampling weights is 14,133,175.86.

There is no known multiplicity within each of the 48 Visit Survey frames (that is, each encounter is believed to have been represented by only one record on each grantee's encounter frame), so no multiplicity adjustment need be made to the encounter-level sampling weight.

b. Encounter-Level Analysis Weight

Like the user-level weight, several adjustments were made to the encounter-level sampling weights to account for non-completion and ineligibility as well as quantifiable undercoverage in two grantees. The types of "nonresponse" for the Visit Survey are patient records that were not located or were unavailable, and so eligibility status is unknown. There were no cases of an encounter known to be eligible, but for which abstraction did not take place. We therefore did the weighting class adjustment for nonresponse in two stages (rather than in three stages, as was done for the User Survey weights): (1) missing encounters

due to grantee nonresponse; and (2) missing encounters due to the unavailability of patient records. A poststratification adjustment similar to that carried out in the User Survey was undertaken as well. The preliminary analysis weights resulting from these adjustments were then examined for outliers. Trimmed weights were considered in cases where the reduction in variance gained by the trimmed weights would outweigh any associated bias introduced in doing so.

Adjustments to the encounter-level sampling weight are discussed, and were implemented, in the following order:

- Weighting class adjustment for encounters missing due to grantee nonresponse
- Poststratification adjustment for two grantees with less than the full year of encounters represented on frames
- Weighting class adjustment for non-located records

Weighting Class Adjustment for Encounters Missing Due to Grantee Nonresponse. The first weighting class adjustment must account for the estimated number of encounters lost due to grantee-level nonresponse (discussed previously under the similar adjustment made to the user sampling weight). Sampling strata were used as the weighting classes for this adjustment, as they were for the comparable user survey adjustment.

For all but the two strata containing refusers, this weighting class adjustment factor is equal to one. For the two strata containing refusers, the weighting class adjustment factor is greater than one. The weighting class adjustment factor for the estimated number of encounters lost to grantee-level nonresponse for stratum *h can* be thought of as the total number of encounters targeted in stratum *h* divided by the total number of encounters represented by the responding grantees in stratum *h*. Let p_h be the ratio of the total number of visits to the total number of users in stratum ***h,*** or:

$$(30)\quad p_h = \frac{\sum_{i=1}^{M_h} \hat{CNOV}_i}{\sum_{i=1}^{M_h} \hat{CMOS}_i}$$

Then the weighting class adjustment factor can be quantified as:

$$(31)\quad A_G(h) = \frac{\sum_{i=1}^{M_h} \hat{CNOV}_i}{\sum_{j=1}^{m_h} W_S(G_j)\, \hat{GMOS}_j\, p_h\, \delta(G_j)} = \frac{m_h}{\sum_{j=1}^{m_h} \delta(G_j)}$$

where $\hat{CNOV}_i$ is an estimate of the number of medical encounters within cluster C,. Although such an estimate for the number of medical encounters is not available for non-selected and non-participating grantees, the terms cancel out, and one is left with the same adjustment factors used for the comparable User Survey adjustment.

An interim encounter-level weight for encounter V_r *can* now be computed as:

$$(32)\quad W_{I_v}(V_r) = W_S(V_r)\, A_G(h)$$

The minimum value of this interim encounter-level weight is 2,488.20 and the maximum value is 6,185.94. The sum of these weights is 14,759,996.73.

Poststratification Adjustment for Two Grantees with Less Than the Full Year of Encounters Represented on Frames. The next step in the encounter weight adjustment process is poststratification. As with the User Survey, the totals being used to calculate these weights (grantee-specified counts of medical encounters) are the best size estimates available, and so we did not poststratify to any sort of external counts. However, two of the grantees were unable to provide frames for the entire 1994 calendar year. It is believed that the encounters from the missed time periods are not meaningfully different than

those on the frame, and so we weighted the encounters from those two grantees so that they represent all their encounters in 1994.

In the User Survey poststratification adjustment, we used a preliminary estimate of users from the grantee, divided by the actual number on the frame, as the adjustment factor. While this was the best estimate of undercoverage available, such preliminary estimates of users from the grantees were often found to be fairly different from actual frame sizes for those providing full-year frames. And although users from exactly half of 1994 were missing from one of the grantees, we could not simply multiply the frame size by two to make the six-month users represent the full-year of users because many users were likely to have used the grantee's services in both halves of the year. This complication does not exist for the Visit Survey component, and so we used the actual frame size for the limited time period and adjusted by the fraction of year that is represented. The encounters from the first grantee had their weights blown up by a factor (PSADJ) of 2, because encounters are missing from January through June. The encounters from the second grantee had their weights blown up by a factor (PSADJ) of 1.5, because encounters are missing from January through April. The factor PSADJ is set equal to one for encounters from all other grantees.

So, an interim encounter-level weight can be calculated as:

(33) $W_{I_v}(V_r) = W_I(V_r) \; PSADJ$

The sum of these weights is 15,031,242.29. The minimum value is 2,488.20 and the maximum value is 7,644.32.

Weighting Class Adjustment for Non-Located Records. The propensity for a medical record to be unavailable for abstraction quite likely is grantee-specific, although this occurred infrequently. The analytically-relevant information collected on the abstraction form is likely to be grantee-specific as well as age- and sex-specific. However, crossing grantee with age group and sex would yield weighting

adjustment classes that are too small. Instead, we defined the weighting adjustment classes by grantee and sex, ensuring that the adjustment factor did not exceed two within any cell and that there were at least 20 completed encounters per cell in any cells with noncompletes[9].

The weighting class adjustment factor for the number of encounters lost because the patient chart was not available or locatable (within class c) can be thought of as the total number of encounters targeted in class c divided by the number of encounters represented by the completed abstractions in class c. Let k_c be the eligibility rate in class c among those with known eligibility status, or:

$$(34)\quad k_c = \frac{\sum_{r' \in E_c} \delta_{elig}(V_{r'})}{\sum_{r' \in E_c} [\delta_{elig}(V_{r'}) + \delta_{inelig}(V_{r'})]}$$

Then the weighting class adjustment factor can be quantified as:

$$(35)\quad A_E(c) = \frac{\sum_{r \in E_c} W_{I_v}(V_r)\,[\delta_{elig}(V_r) + k_c\,\delta_{undet}(V_r)]}{\sum_{r \in E_c} W_{I_v}(V_r)\delta_{elig}(V_r)\delta_{comp}(V_r)}$$

where:

E_c	is the set of all selected and released encounters in class c
$\delta_{elig}(V_r)$	is an indicator variable that is equal to 1 when encounter V_r is known to be eligible, 0 otherwise
$\delta_{inelig}(V_r)$	is an indicator variable that is equal to 1 when encounter V_r is known to be ineligible, 0 otherwise
$\delta_{undet}(V_r)$	is an indicator variable that is equal to 1 when encounter V_r's eligibility status is undetermined, 0 otherwise $[\delta_{undet} = 1 - (\delta_{elig} + \delta_{inelig})]$
$\delta_{comp}(V_r)$	is an indicator variable that is equal to 1 when the abstraction for encounter V_r is complete, 0 otherwise.

'There were four cells (out of 90) that had between 16 and 19 completes per cell, with adjustment factors ranging from 1.05263 to 1.11765.

In combining the two weighting class adjustment factors for nonresponse with the encounter sampling weight, the final analysis weight (aside from any trimming) for encounter V_r at grantee G_j and also in class c is:

$$(36)\quad W_A(V_r) = W_{I'_v}(V_r)\, A_E(c)$$

Among those with nonzero weights, this weight had a minimum value of 2,488.20 and a maximum value of 7,644.32. (There were 62 cases--those with no chart available--with zero weights.) The sum of these weights is equal to 15,031,242.29. Among the 2,878 completes, the weights ranged from 2,488.20 to 7,644.32, with a sum of 13,893,454.69.

c. Truncation

There was very little heterogeneity among the weights for the visit abstraction. The design effect due to unequal weighting was only 1.03. We therefore concluded that there was little to be gained by trimming the weights. Nevertheless, we explored the impact on individual estimates of trimming the weights. The largest decrease in variance from trimming at the 10th and 90th percentiles was found in "provider seen-other" (standard error reduced from 15.05 to 13.99), followed by "disposition-return appointment" (11.29 to 10.47) Hispanic ethnicity (37.21 to 36.37) and "race-other" (15.92 to 15.44). Trimming the weights at the 10th and 90th percentiles had little impact on the estimates. The estimates affected most by such trimming were percent black (3 1.85 percent untruncated versus 32.19 percent truncated), percent seen before for same condition (49.00 percent versus 49.35 percent), and percent with disposition of "return as needed' (28.28 percent versus 27.91 percent).

d. Variance and Design Effects

As described above for the user survey, the variance of the estimates from the encounter abstraction were calculated using SUDAAN software, using the with-replacement design option. Table II. 18 presents

TABLE II. 18

ESTIMATES AND STANDARD ERRORS FOR ENCOUNTER ABSTRACTIONS

Variable Number and Name		Number Responding	Unweighted Frequencies	Weighted Estimate (percent*)	Standard Error (SUDAAN)
Q2_YR	Year of birth (19xx)	2878		63.71	1.43
Q3	Sex - Female	2878	1851	64.09	0.94
Q4	Race - White	2394	1158	48.65	4.93
	Race - Black		771	31.85	4.45
	Race - Asian/Pacific Isl		121	5.20	3.79
	Race - Am. Indian		12	0.48	0.31
	Race - Other		332	13.82	3.99
Q5	Hispanic ethnicity	1904	803	41.75	6.10
Q6	Reimb. - Mgd Care/HMO	2795	468	16.70	2.55
	Reimb. - Other		2137	76.76	3.42
	Reimb. - Info. not avail.		190	6.54	2.29
Q7_1	Source of pmt. -Medicare	2861	350	12.99	2.17
Q7_2	Source of pmt.-Medicaid		1271	44.29	3.21
Q7_3	Source of pmt.-oth. govt.		183	6.08	1.97
Q7_4	Source of pmt.-private		448	16.48	2.14
Q7_5	Source of pmt.-patient		901	31.45	2.93
Q7_6	Source of pmt.-no charge		72	2.59	1.16
Q7_7	Source of pmt.-other		101	3.41	0.81
Q8	Referred by this center	2683	204	7.53	2.05
	Referred by outside ctr		47	1.74	0.32
	Not referred		2432	90.74	2.20
Q11A	Seen in CHC before	2762	2445	88.78	1.09
Q11B	For same condition	2353	1173	49.00	2.77
Q13_1	Disp. -Return PRN	2860	796	28.28	3.09
Q13_2	Disp. -Return appt.		1698	58.98	3.36
Q13_3	Disp.-Telephone f-up		36	1.28	0.34
Q13_4	Disp.-Referring physician		50	1.73	0.82
Q13_5	Disp.-Other physcn/clinic		162	5.69	0.65
Q13_6	Disp. -Admit to hospital		13	0.44	0.14
Q13_7	Disp. -No follow-up		331	11.56	2.11
Q13_8	Disp.-Other		120	4.20	0.84
Q14_1	Saw physician	2859	2297	80.26	2.40

TABLE II. 18 *(continued)*

Variable Number and Name		Number Responding	Unweighted Frequencies	Weighted Estimate (percent*)	Standard Error (SUDAAN)
Q14_2	Saw physician asst		408	14.74	3.28
Q14_3	Saw nurse practitioner		350	11.43	2.32
Q14_4	Saw nurse-midwife		91	3.09	1.29
Q14_5	Saw registered nurse		236	7.89	1.80
Q14_6	Saw other provider		343	12.11	3.88
Q15_1	Blood pressure	2856	1785	61.89	3.18
Q15_2	Urinalysis		429	15.14	1.32
Q15_3	EKG resting		31	1.09	0.21
Q15_4	EKG exercise		4	0.16	0.10
Q15_5	Mammogram		21	0.77	0.17
Q15_6	Chest x-ray		53	1.80	0.26
Q15_7	Other radiology		56	2.02	0.39
Q15_8	Allergy testing		3	0.12	0.06
Q15_9	Spirometry/lung test		29	1.11	0.82
Q15_10	Pap test		124	4.33	0.41
Q15_11	Strep throat test		60	2.05	0.43
Q15_12	HIV serology		23	0.81	0.21
Q15_13	Other blood test		438	15.54	1.42
Q15_14	Other lab test		242	8.43	1.19
Q15_15	Hearing test		61	2.20	0.54
Q15_16	Visual acuity		97	3.45	0.65
Q15_17	Mental status exam		1	0.03	0.03
Q15_18	Surgical procedure		26	0.93	0.23
Q15_19	Other procedure (1)		389	13.82	3.23
Q15_20	Other procedure (2)		64	2.24	0.40
Q16_1	Diet/nutritional counseling	2826	305	10.66	1.06
Q16_2	Exercise		99	3.42	0.75
Q16_3	Cholesterol reduction		27	1.02	0.19
Q16_4	Weight reduction		27	0.94	0.21
Q16_5	Smoking cessation		40	1.34	0.28
Q16_6	Family planning		76	2.62	0.43
Q16_7	Prenatal/parenting class		51	1.74	0.38
Q16_8	Growth/development		88	3.16	1.12
Q16_9	Injury prevention		39	1.33	0.29
Q16_10	HIV transmission		45	1.57	0.34
Q16_11	Other STD transmission		47	1.68	0.29

TABLE II. 18 ***(continued)***

Variable Number and Name		Number Responding	Unweighted Frequencies	Weighted Estimate (percent*)	Standard Error (SUDAAN)
Q16_12	Physiotherapy		13	0.48	0.24
Q16_13	Psychotherapy/MH trmt		15	0.54	0.13
Q16_14	Family Counseling		22	0.78	0.29
Q16_15	Drug abuse trmt/cnsling		10	0.34	0.12
Q16_16	Alcohol abuse trmt/cnslg		10	0.34	0.15
Q16_17	Occup/vocatl counseling		5	0.19	0.15
Q16_18	Educ/literacy training		5	0.18	0.08
Q16_19	WIC services/counseling		24	0.81	0.21
Q16_20	Case management		10	0.34	0.15
Q16_21	Eligibility assistance		7	0.24	0.10
Q16_22	Corrective lenses		8	0.27	0.13
Q16_23	Hearing aid		1	0.04	0.04
Q16_24	Domestic violence prev.		2	0.07	0.07
Q16_25	Environmtl hlth risk reduc.		6	0.22	0.12
Q16_26	Housing assistance		1	0.04	0.04
Q16_27	Interpretation/translation		6	0.20	0.09
Q16_28	Dental care-preventive		29	1.02	0.28
Q16_29	Dental care-restorative		6	0.22	0.10
Q16_30	Other services		191	6.85	0.91
Q17	Second med. encounter	2794	11	0.38	0.14

* Year of birth estimate is a mean, not percent.

estimates and standard errors for the encounter survey. Table II. 19 shows the design effects for all estimates from the visit abstraction (except for those variables describing the second encounter, which only pertain to eleven cases out of the 2,878 completes).

The design effects are quite large for most of the estimates, due almost entirely to the effect of the clustered design. As stated above, the component of the design effect due to unequal weighting was only 1.03. Achieving such a small unequal weighting component of the design effect is the result of our efforts during the sampling process to equalize probabilities of selection across all grantees, to the extent possible. The clustering effect was high because most of the variables being estimated are related to grantee characteristics, and the grantee was the primary sampling unit. Because a clustering effect decreases as the number of primary sampling units increases, a larger number of grantees with the same number of observations would have increased the effective sample size. Likewise, because a clustering effect increases as the number of observations within each cluster increases, augmenting the number of observations within the same number of grantees would have increased the clustering effect and only marginally increased the precision of the estimates.

Especially high are the estimated design effects for race and Hispanic ethnicity (29.83 and 28.5 1, respectively), followed by provider type (19.87 averaged over all categories), reimbursement type (18.43, averaged), source of payment (12.18, averaged), referral status (11.07, averaged), and year of birth[10] (10.36). The smallest design effects were for sex (1.1 1), whether there was a second medical encounter (1.45), and types of services received (2.32, averaged). These design effects can be compared to an estimate calculated by NCHS when developing the initial sample design. At that time a design effect of 10 was estimated for visit survey planning purposes.

'When looking at proportions in four different age categories, the design effects were smaller: ages O-17, 5.86; ages 18-34, 1.15; ages 35-64, 3.26; and ages 65 and older, 9.38.

TABLE II. 19

DESIGN EFFECTS FOR UNTRUNCATED WEIGHTS

Variable Number and Name		Design Effect By Estimate	Design Effect By Question	Effective Sample Size (nominal n=2,878)
Q2_YR	Year of birth	10.36	10.36	277.80
Q3	Sex	1.11	1.11	2592.79
Q4	Race - White	22.94	29.83	96.49
	Race - Black	21.50		
	Race - Asian/Pacific Isl	68.60		
	Race - Am. Indian	4.63		
	Race - Other	31.46		
Q5	Hispanic ethnicity	28.51	28.51	100.95
Q6	Reimb. - Mgd Care/HMO	13.02	18.43	156.13
	Reimb. - Other	18.33		
	Reimb. - Info. not avail.	23.95		
Q7_1	Source of pmt.-Medicare	11.88	12.18	236.32
Q7_2	Source of pmt.-Medicaid	11.91		
Q7_3	Source of pmt.-0th. govt.	19.46		
Q7_4	Source of pmt.-private	9.54		
Q7_5	Source of pmt.-patient	11.41		
Q7_6	Source of pmt.-no charge	15.39		
Q7_7	Source of pmt.-other	5.66		
Q8	Referred by this center	16.20	11.07	259.90
	Referred by outside ctr	1.61		
	Not referred	15.41		
Q11A	Seen in CHC before	3.27	3.27	880.12
Q11B	For same condition	7.22	7.22	398.62
Q13_1	Disp. -Return PRN	13.42	7.73	372.32
Q13_2	Disp.-Return appt.	13.34		
Q13_3	Disp.-Telephone f-up	2.64		
Q13_4	Disp.-Referring physician	11.41		
Q13_5	Disp.-Other physcn/clinic	2.27		
Q13_6	Disp.-Admit to hospital	1.24		
Q13_7	Disp.-No follow-up	12.49		
Q13_8	Disp.-Other	5.03		
Q14_1	Saw physician	10.40	19.87	144.84
Q14_2	Saw physician asst	24.54		
Q14_3	Saw nurse practitioner	15.13		

TABLE II. 19 (continued)

Variable Number and Name		Design Effect By Estimate	Design Effect By Question	Effective Sample Size (nominal n=2,878)
Q14_4	Saw nurse-midwife	15.85		
Q14_5	Saw registered nurse	12.81		
Q14_6	Saw other provider	40.49		
Q15_1	Blood pressure	12.26	4.69	613.12
Q15_2	Urinalysis	3.85		
Q15_3	EKG resting	1.16		
Q15_4	EKG exercise	1.72		
Q15_5	Mammogram	1.03		
Q15_6	Chest x-ray	1.10		
Q15_7	Other radiology	2.14		
Q15_8	Allergy testing	1.00		
Q15_9	Spirometry/lung test	17.54		
Q15_10	Pap test	1.15		
Q15_11	Strep throat test	2.57		
Q15_12	HIV serology	1.59		
Q15_13	Other blood test	4.40		
Q15_14	Other lab test	5.21		
Q15_15	Hearing test	3.90		
Q15_16	Visual acuity	3.66		
Q15_17	Mental status exam	0.95		
Q15_18	Surgical procedure	1.62		
Q15_19	Other procedure (1)	24.98		
Q15_20	Other procedure (2)	2.05		
Q16_1	Diet/nutritional counseling	3.36	2.32	1241.23
Q16_2	Exercise	4.81		
Q16_3	Cholesterol reduction	1.04		
Q16_4	Weight reduction	1.28		
Q16_5	Smoking cessation	1.65		
Q16_6	Family planning	2.00		
Q16_7	Prenatal/parenting class	2.40		
Q16_8	Growth/development	11.69		
Q16_9	Injury prevention	1.77		
Q16_10	HIV transmission	2.16		
Q16_11	Other STD transmission	1.39		
Q16_12	Physiotherapy	3.55		
Q16_13	Psychotherapy/MH trmt	0.93		

TABLE II. 19 *(continued)*

Variable Number and Name		Design Effect By Estimate	Design Effect By Question	Effective Sample Size (nominal n=2,878)
Q16_14	Family Counseling	3.19		
Q16_15	Drug abuse trmt/cnsling	1.27		
Q16_16	Alcohol abuse trmt/cnslg	1.93		
Q16_17	Occup/vocatl counseling	3.46		
Q16_18	Educ/literacy training	1.03		
Q16_19	WIC services/counseling	1.52		
Q16_20	Case management	1.85		
Q16_21	Eligibility assistance	1.15		
Q16_22	Corrective lenses	1.66		
Q16_23	Hearing aid	1.03		
Q16_24	Domestic violence prev.	2.07		
Q16_25	Environmtl hlth risk reduc.	1.97		
Q16_26	Housing assistance	1.13		
Q16_27	Interpretation/translation	1.07		
Q16_28	Dental care-preventive	2.21		
Q16_29	Dental care-restorative	1.35		
Q16_30	Other services	3.64		
Q17	Second med. encounter	1.45	1.45	1984.83

5. Special Coding and Editing

Data from the NHAMCS are coded by trained medical coders from the NCHS computer facility at Research Triangle Park, North Carolina. The CHC data were coded by an experienced coder who had recently retired from that facility. Her work was reviewed by MPR staff, and some cases were returned to the coder for review and modification. The use of a coder trained in NHAMCS coding procedures helped ensure comparability between the two databases. The patient's complaint, symptom, or other reason for visit was coded according to Reason for Visit Classification and Coding Manual." Up to seven were coded. The physician's diagnosis was coded or the existing ICD-9 codes were checked according to the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM). Up to seven were coded. The medication data were coded according to a scheme developed by NCHS based on the American Society of Hospital Pharmacists' Drug Product Information File, which is maintained by the American Druggist Blue Book Data Center. Up to 11 medications were coded.

Data from the medical encounter abstraction were entered into three different data files: (1) the abstraction data, (2) the "specify" data, and (3) the coding data. These three files were then merged together into one data file. The "specify" data were cleaned so that each time the "other" response was checked, there was "specify" data entered, and vice versa. The coding data were verified, as described previously, and updated. We checked that each patient complaint, physician diagnosis, and medication had a corresponding code or codes.

We first checked for response and eligibility status. To check for response status, we not only checked that the form had been completed, but we checked that it had been completed for the sampled medical encounter. If the date of birth, sex, or date of visit did not match what was on the sample information given to us, then we called the grantee to adjudicate. This resulted in four abstractions having to be re-done. To

---

"National Center for Health Statistics, CDC/PHS, U.S. Department of Health and Human Services (1994), Hyattsville, MD.

check for eligibility status, we checked that the visit took place in calendar year 1994 and that a physician or mid-level practitioner was seen. This required calls back to the grantees as well, particularly when the "other" provider had been checked and it was not clear whether this provider was eligible. This check resulted in our losing 45 cases: 34 cases because no eligible provider was seen; six cases because the visit took place after 1994; four cases because they were no-shows; and one case because the selected encounter was in a hospital, rather than the CHC. Checking for response and eligibility status also generated edits for final disposition of case, date of birth, sex, date of visit, and provider seen.

The remainder of the items on the abstraction form were also edited, with missing values being coded as logical skips, "don't know," or data missing, as appropriate. We also cleaned up the code for site of encounter, and checked whether the third page of the form, which was for a second medical encounter the same day as the sampled encounter, was complete and whether it was an eligible medical encounter.

REFERENCES

American Medical Association. *Physicians Current Procedural Terminology Manual.* Chicago: AMA, 1994.

Bureau of Primary Health Center, Health Resources and Services Administration, Public Health Service. *The Bureau Common Reporting Requirements Manual.* Washington, DC: U.S. Department of Health and Human Services, 1991.

Office of Management and Budget. *Standard Industrial Classification Manual.* Washington, DC: OMB, 1987.

U.S. Department of Health and Human Services. *Public Use Data Tape Documentation, Part Three: Medical Coding Manual and Short Index.* Washington, DC: HHS, 1988.

U.S. Department of Commerce, Office of Federal Statistical Policy and Standards. *Standard Occupational Classification Manual.* Washington, DC: U.S. Department of Commerce, 1980.

APPENDIX 1

FORMULAS FOR SELECTION OF USER AND ENCOUNTER SAMPLES

APPENDIX 1

Initial formula for the number of users to select per grantee, for granteej with no site subsampling:

$$u_j = \frac{40}{PCR} \cdot \frac{50\ GMOS_j}{G\hat{M}OS_j \sum_{g=1}^{49} \frac{GM\text{-}OS,}{G\hat{M}OS_g}}$$

where PCR is the initial predicted completion rate for the user survey (product of eligibility rate and response rate among eligibles). PCR was initially set to (.95)(.76)=.722. (Other elements in the equation are defined in the text.)

Initial formula for the number of users to select per site, for site *k* in granteej (with site subsampling):

$$u_k = \frac{1}{s_j} \cdot \frac{40}{PCR} \cdot \frac{\mathbf{50}\ G\tilde{M}OS_j\ SMOS_k}{G\hat{M}OS_j\ S\tilde{M}OS_k \sum_{g=1}^{49} \frac{G\tilde{M}OS_g}{G\hat{M}OS_g}}$$

Final formula for the number of users to select per grantee, for grantee *j* with no site subsampling, in stratum *h*:

$$u_j = \frac{41.6}{PCR'} \cdot \frac{GMOS_j \sum_{c-1}^{y_h} C\hat{M}OS_c}{A\ .\ G\hat{M}OS_j\ .\ x_h}$$

where $A = \frac{1}{48}\left[\sum_{g=1}^{G_h} \frac{GMOS_g \sum_{c=1}^{y_h} CM\hat{O}S_c}{GM\hat{O}S_g \cdot x_h} + \sum_{g=1}^{G_h}\sum_{s=1}^{s_g} \frac{GM\tilde{O}S_g \cdot SMOS_s \cdot \sum_{c=1}^{y_h} CM\hat{O}S_c}{GM\hat{O}S_g \cdot SM\tilde{O}S_s \cdot s_g \cdot x_h}\right]$

and PCR' is the revised predicted completion rate for the user survey (product of eligibility rate and response rate among eligibles). PCR' was estimated to be (.90)(.76)=.684. The first summation within the square brackets is over those grantees with no site subsampling. The second summation within the brackets is over those sites subsampled from all other grantees. (Other elements in the equation are defined in the text.)

Final formula for the number of users to select per site, for site *k in* granteej (with site subsampling), in stratum *h:*

$$u_k = \frac{1}{s_j} \cdot \frac{41.\bar{6}}{PCR'} \cdot \frac{GM\tilde{O}S_j \cdot SMOS_k \cdot \sum_{c=1}^{y_h} CM\hat{O}S_c}{A \cdot GM\hat{O}S_j \cdot SM\tilde{O}S_k \cdot x_h}$$

Initial formula for the number of encounters to select per grantee, for granteej in stratum *h:*

$$v_j = \frac{60}{PCR''} \cdot \frac{GNOV_j \sum_{c=1}^{y_h} CM\hat{O}S_c}{B \cdot GM\hat{O}S_j \cdot x_h}$$

where $B = \frac{1}{50}\sum_{h=1}^{11}\sum_{g=1}^{x_h'} \frac{GN\tilde{O}V_g}{GM\hat{O}S_g} \cdot \frac{\sum_{c=1}^{y_h} CM\hat{O}S_c}{x_h}$

and PCR" is the predicted completion rate for the abstraction effort (product of eligibility rate and response rate among eligibles). PCR" was estimated to be (.99)(.95)=.9405. The term x_h' refers to the number of clusters with participating grantees in stratum *h*. (Other elements in the equation are defined in the text.)

Final formula for the number of encounters to select per grantee, for granteej in stratum h:

$$v_j = \frac{60}{PCR''} \cdot \frac{GNOV_j \sum_{c=1}^{y_h} \hat{CMOS}_c}{B' \cdot \hat{GMOS}_j \cdot x_h}$$

where $B' = \frac{1}{50} \sum_{h=1}^{11} \sum_{g=1}^{x_h'} \frac{GNOV_g}{\hat{GMOS}_g} \cdot \frac{\sum_{c=1}^{y_h} \hat{CMOS}_c}{x_h}$

APPENDIX 2

MEDICAL ENCOUNTER ABSTRACT FORM

OMB No.: 0915-0185
Exp: 3/31/96

NATIONAL STUDY OF COMMUNITY HEALTH CENTERS:

MEDICAL ENCOUNTER ABSTRACT FORM

Department of Health and Human Services

U.S. Public Health Service
Health Resources and Services Administration

GENERAL INSTRUCTIONS

Questions 1 through 14 refer to the sampled medical encounter. Questions 15 and 16 concern other services or procedures **provided on the same date as the sampled medical encounter. Question 17 determines whether there is a second medical encounter on that date. Questions 18 through 24 refer to the second medical encounter.**

If you have any questions, please contact:

Rhoda Cohen
Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ 08543-2393

1-800-232-8024
(609) 275-2324

LABEL

1. SITE OF ENCOUNTER *(Record name and # of site from manual if available)*

2. DATE OF BIRTH

____/____/____
Month Day Year

3. SEX

1 ☐ Female
1 ☐ Male

4. RACE

☐ Don't Know
1 ☐ White
2 ☐ Black
3 ☐ Asian/Pacific Islander
4 ☐ American Indian/Eskimo/Aleut
5 ☐ Other

6. ETHNICITY

1 ☐ Hispanic
2 ☐ Not Hispanic
☐ Don't Know

6. REIMBURSEMENT TYPE *(for sampled medical encounter/*

1 ☐ Managed care/HMO/Prepaid
2 ☐ Other
3 0 Information not available

7. EXPECTED SOURCE(S) OF PAYMENT *(Check all that apply for sampled medical encounter)*

1 ☐ Medicare
2 ☐ Medicaid
3 ☐ Other government
4 ☐ Private/Commercial
5 ☐ Patient paid
6 ☐ No charge
7 ☐ Other *(Specify)*

8. *WAS* PATIENT REFERRED FOR SAMPLED MEDICAL ENCOUNTER BY ANOTHER MEDICAL PROVIDER?

1 ☐ Yes, this center
2 ☐ Yes, outside center
3 ☐ No
☐ Don't Know

9. PATIENT'S **COMPLAINT(S), SYMPTOM(S),** OR OTHER **REASON(S)** FOR SAMPLED MEDICAL ENCOUNTER *(in patient's own words)*

a. Most important:

b. Other:

c. Other:

10. PHYSICIAN'S **DIAGNOSES** FOR SAMPLED MEDICAL ENCOUNTER

Enter ICD-9

a. Principal diagnosis/problem associated with item 9a:

b. Other:

c. Other:

116. HAS PATIENT BEEN SEEN IN THIS COMMUNITY HEALTH CENTER BEFORE? *(At any center or site within grantee system)*

1 ☐ Y66-600TO IIb
2 ☐ No → GO TO 12

IIb. WAS THIS FOR **CONDITION** RECORDED IN ITEM 10, PART a?

1 ☐ Yes
2 ☐ No
0 Don't Know

12. MEDICATIONS/INJECTIONS (ASSOCIATED WITH SAMPLED MEDICAL ENCOUNTER)

Include:
- Rx *end OTC*
- *Immunizations*
- *Allergy Shots*
- *Anesthetics*
- *Meds ordered, supplied, or administered*
- *New meds*
- *Continuing meds--specified*
- *Continuing meds--unspoecified ("continue present regimen")*

☐ None

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
a. _____
10. _____

13. DISPOSITION OF SAMPLED MEDICAL ENCOUNTER *(Check all that apply)*

1 ☐ Return to clinic PRN
2 0 Return to clinic appointment
3 3 Telephone follow-up planned
4 ☐ *Return to* referring physician
5 3 Refer to other physician/clinic
6 0 Admit to hospital
7 ☐ No follow-up planned
8 ☐ Other *(Specify)* _____

14. PROVIDERS SEEN FOR SAMPLED MEDICAL ENCOUNTER *(Check all that apply)*

1 ☐ Physician
2 ☐ Physician assistant
3 0 Nurse practitioner
4 ☐ Nurse-Midwife
5 0 Registered nurse
6 0 Other _____

16. DIAGNOSTIC/SCREENING SERVICES, SURGICAL AND NONSURGICAL PROCEDURES *(Check all provided on the date of sampled medical encounter/*

0 None

1 ☐ Blood pressure
2 0 Urinalysis
3 0 EKG resting
4 0 EKG exercise
5 0 Mammogram
6 *0* Chest x-ray
7 0 Other radiology
8 0 Allergy testing
9 0 Spirometry/Other lung-function test
10 0 Psp test
11 0 Strep throat test
12 0 HIV serology
13 0 Other blood test
14 0 Other lab test
15 ☐ Hearing test
16 0 Visual acuity
17 0 Mental status exam
18 ☐ Surgical procedures *(Specify)* _____
19 0 Other *(Specify)* _____
20 0 Other *(Specify)* _____

16. COUNSELING/EDUCATION AND OTHER SERVICES
(Check all ordered or provided on ths date of sampled medical encounter: Exclude medication)

☐ *None*

Counseling/Education

1 0 Diet/Nutritional Counseling
2 ☐ Exercise
3 0 Cholesterol Reduction
4 0 Weight Reduction
5 0 Smoking Cessation
6 ☐ Family Planning
7 0 Prenatal/Parenting Class
8 0 Growth/Development
9 0 Injury Prevention
10 *0* HIV Transmission
11 ☐ Other STD Transmission

Selected Services

12 0 Physiotherapy
13 0 Psychotherapy/Mental Health Treatment
14 0 Family Counseling
15 0 Drug Abuse Treatment/Counseling
16 0 Alcohol Abuse Treatment/Counseling
17 0 Occupational/Vocational Counseling
18 0 Educational/Literacy Training
19 0 WIC Services/Counseling
20 0 Case Management
21 0 Eligibility Assistance

All Other Services

22 0 Corrective Lenses
23 ☐ Hearing Aid
24 0 Domestic Violence Prevention
25 0 Environmental Health Riik Reduction
26 0 Housing Assistance
27 0 Interpretation/Translation
28 ☐ Dental Care Preventive
29 0 Dental Care-Restorative
30 0 Other *(Specify)* _____

17. IS THERE DOCUMENTATION IN THE MEDICAL RECORD FOR ANOTHER MEDICAL ENCOUNTER ON THE SAME DATE AS THE SAMPLED ENCOUNTER7
(See label on page1 for date of sampled encounter.)

1 0 Yss ➔ COMPLETE ITEMS 18-24 ON NEXT PAGE

2 0 No ➔ THANK YOU FOR COMPLETING THIS FORM

The following questions refer to a second medical encounter on the same date as sampled encounter.

18. WAS PATIENT REFERRED FOR THE SECOND MEDICAL ENCOUNTER BY ANOTHER MEDICAL PROVIDER?

1 ☐ Yes, this center
2 ☐ Yes, outside center
3 ☐ No
☐ Don't Know

IS. PATIENT'S COMPLAINT(S), SYMPTOM(S), OR OTHER REASON(S) FOR SECOND MEDICAL ENCOUNTER *(in patient's own words)*

a. Most important:

b. Other:

c. Other:

20. PHYSICIAN'S DIAGNOSES (FOR SECOND MEDICAL ENCOUNTER)

ICD-9

a. Principal diagnosis/problem associated with item 19a:

b. Other:

c. Other:

21a. HAS PATIENT BEEN SEEN IN THIS COMMUNITY HEALTH CENTER BEFORE? *(At any center or site within grantee system/*

1 ☐ Yes → GO TO 21b
2 ☐ No → GO TO 22

21b. WAS THIS FOR CONDITION RECORDED IN ITEM 20. PART *a7*

1 ☐ *Yes*
2 ☐ *No*
☐ Don't Know

22. MEDICATIONS/INJECTIONS (ASSOCIATED WITH SECOND MEDICAL ENCOUNTER)

Include:
- *Rx and* OTC
- *Immunizations*
- *Allergy Shots*
- *Anesthetics*
- *Meds ordered, supplied, or administered*
- *New meds*
- *Continuing mods-specified*
- *Continuing meds--unspecified ("continue present regimen")*

☐ *None*

1. 6.
2. 7.
3. 8.
4. 9.
5. 10.

23. **DISPOSITION OF SECOND MEDICAL ENCOUNTER** *(Check all that apply)*

1 ☐ Return to clinic PRN
2 ☐ Return to clinic appointment
3 3 Telephone follow-up planned
4 ☐ Return to referring physician
5 ☐ Refer to other physician/clinic
6 ☐ Admit to hospital
7 ☐ No follow-up planned
8 ☐ Other *(Specify)*

24. PROVIDERS SEEN FOR SECOND MEDICAL ENCOUNTER *(Check all that apply)*

1 *0* Physician
2 ☐ Physician assistant
3 ☐ Nurse practitioner
4 ☐ Nurse-Midwife
5 ☐ Registered nurse
6 ☐ Other

APPENDIX 3

USER SAMPLING INSTRUCTIONS

SAMPLE SELECTION FOR THE USER SURVEY

This document will provide you with instructions for selecting a sample of community health center (CHC) users. The intention of the User Survey is to study patients who used CHCs for medical care in 1994. Information for the User component of the study will be gathered through in-person interviews. We expect to collect data on about 40 patients per CHC.

We define a ***CHC*** or grantee as the administrative entity that receives 330 funding from the Bureau of Primary Health Care in the Health Resources and Services Administration. A CHC or grantee can have one or more ***sites***. A site can be thought of as a single location or facility. Only those permanent sites receiving 330 funding ***and*** providing general medical services are eligible for this study. Sites that are 100% migrant worker or 100% homeless are ineligible. We have selected between one and three sites (among all of your eligible sites) for inclusion in the study. Your field representative will tell you which sites have been selected.

We define a ***user as*** someone who has made at least one medical visit to a selected site for care in 1994. We define a ***medical visit as*** a visit to a selected site in 1994 where a physician, midlevel practitioner (nurse practitioner, nurse-midwife, physician's assistant), or nurse is seen. Dental and other health visits are not considered medical visits, nor are services such as laboratory, x-ray, and prescriptions, unless a physician, midlevel practitioner, or nurse was also seen.

Certain users are ineligible for the User Survey, primarily due to the difficulty with which they could be contacted. They include migrant workers, homeless persons, and persons who have died since their last CHC visit. ***Migrant workers are*** defined in the same way that you report them to the Bureau. Note that seasonal farm workers are eligible for this study. The ***homeless are*** defined as persons who either: have no address listed in the medical record (no way to contact them) or are reported by family members, friends, or someone else to be homeless at the time of the survey. For those CHCs with selected school-based clinic sites, students who have parental permission only for limited services and only at the school-based site should be considered ineligible as well. Anyone known by the CHC to have moved out of the CHC's service area since his or her last visit or for whom it is known that the person will be out of the service area for the entire period of data collection (April through

<sub>

July of 1995) is also to be considered ineligible.

To obtain a representative sample of users from each CHC, it is critical that these sampling instructions be followed. Different CHCs may require slightly different procedures, depending on the level of automation of records. All acceptable procedures are described below. After you have read the instructions, please call your field representative. Any questions you have can be addressed at that time. We understand that some of what we are asking may be difficult to carry out for your CHC. If you predict or encounter any problem in implementing these procedures, do not attempt an alternative solution on your own. Discuss them with your field representative, who may refer you to the project's sampling manager or one of the project's programmers, depending on the nature of the problem. The best sampling plan for your particular CHC will be determined.

The procedures for selecting the user sample are similar to those for the Visit component, but each must be carried out separately. More detailed instructions follow, but the major steps are:

1. Read the instructions carefully and call your field representative
2. Compile the user list (for each selected site)
3. Exclude ineligible users and duplicates
4. Call your field representative and give list totals
5. Sort each list according to instructions and system capabilities
6. Number each list consecutively
7. Field representative will call you with the information necessary to select the samples
8. Select the user samples
9. Send lists of selected samples to your field representative

In CHCs with computer files of users, the sampling specifications should be relatively easy to accomplish. In most instances, these instructions can also be completed manually using hardcopy lists without too much difficulty.

<sub>

Compile the user list (for each selected site)

Selection of a user sample is straightforward once the list of eligible users has been established. As stated above, the eligible user list must include **all** persons who made at least one medical visit to each **selected** site for care during calendar year 1994. The eligible list must be compiled from all available lists of users, whether they be computerized or hard copy lists. A separate list should be created for each selected site. Be sure you include all users, including those with particularly confidential records (e.g., HIV patients) and those who used specialty services, where separate records may be maintained. ***Record the total number of users for each selected site on the attached worksheet.***

Exclude ineligible users and duplicates

To the extent possible, you must also be sure that each compiled list contains no more than one record per user, for those who may be listed in more than one place, particularly if you have more than one selected site. Each site's list should be checked against the other(s) for duplicates. ***Mark down on the worksheet if you cannot do this.***

In addition, as discussed above, certain types of users are ineligible. Some of these can be determined before the sample is selected, while others will be not be discovered until an attempt is made to contact the person. In creating the lists and selecting the samples, you only need to be concerned about the eligibility that can be determined from your system at the time the list is developed. The following types of users are ineligible (and were defined above): users who made medical visits only to non-selected sites; migrant workers; homeless; deceased; moved out of service area; out of area through data collection period; and students from school-based clinic sites with limited parental permission. Remove all duplicate and ineligible users from each list and ***record for each selected site the total number of users excluded as duplicates and the total number of users excluded as ineligible on the worksheet.***

Call your field representative and give list totals

For each selected site, determine an exact count of all users in the eligible list, with duplicates and ineligible users excluded. ***Record these numbers on the worksheet.*** Then call your field representative. The numbers will be forwarded to the sampling manager, who will determine which records you are to select.

<sub>

Sort each list according to instructions and system capabilities

If your system can do it, sort each site's user list by managed care status, then sex (within managed care status), then prenatal status (within sex). A user should be considered ***prenatal*** is she had at least one prenatal visit in 1994. If some but not all of these items are available for sorting, just sort by what you can. ***Record on the worksheet what you were able to sort the user lists by and what order the user records were in originally. If ordered by medical record number, what does the order of the record number imply?***

Example of Resulting Sort Order:

Managed care	*Male*	*Not prenatal*
Managed care	*Female*	*Prenatal*
Managed care	*Female*	*Not prenatal*
Not managed care	*Male*	*Not prenatal*
Not managed care	*Female*	*Prenatal*
Not managed care	*Female*	*Not prenatal*

Number each list consecutively

If your user lists are on the computer, assign a 6-digit number to each user on each site's eligible list in the sorted order, beginning with the number 000001. If your list is on paper, you can manually number every tenth entry by tens. Check ***that the last number on each user list matches the number previously recorded on the worksheet and reported to your field representative.***

Field representative will call you with the information necessary to select the samples

Your field representative will call you to tell you the which records to select. You will get a different set of numbers for each selected site.

Select the user samples

Select the user sample for each selected site by extracting the records corresponding to the numbers given to you by your field representative. ***Record the number of users you have selected on your worksheet.***

<sub>

Produce each site's user sample list on paper, including all sampled users in the order selected. The list should contain as much of the following information as is available from the files: user's assigned number, complete name, address, telephone number(s), date of birth, sex, medical record number, principal service site used, parent's name (if under 18), and date of last CHC visit. In addition, for users with special confidentiality concerns (HIV patients, emancipated minors), alternative contact information and instructions should be obtained.

Send lists of selected samples to your field representative

Send the lists described above to your field representative as soon possible. Fax the paper lists. *Record on your worksheet when you sent the lists to your field representative.* Fax your worksheet to your field representative along with the sample lists.

<sub>

USER SAMPLE LIST

Name of CHC (grantee)	
BCRR ID#	
Person to contact in case of questions	

PLEASE PRINT ALL INFORMATION FOR EACH OF THE SELECTED USERS.

Eligible list number: |__|__|__|__|__|__|

Complete name: _____
(First Name) (Last Name)

Date of birth: |__|__|/|__|__|/|__|__| Sex: 0 ☐ Male 1 ☐ Female

Contact address: _____

City: _____

State: |__|__| ZIP: |__|__|__|__|__|

Contact telephone numbers: (____) _____ (____) _____

Parents name (if under 18): _____

Date of last visit to CHC: |__|__|/|__|__|/|__|__|

Medical record number: _____

Principal service site used: _____

Eligible list number: |__|__|__|__|__|__|

Complete name: _____
(First Name) (Last Name)

Date of birth: |__|__|/|__|__|/|__|__| Sex: 0 ☐ Male 1 ☐ Female

Contact address: _____

City: _____

State: |__|__| ZIP: |__|__|__|__|__|

Contact telephone numbers: (____) _____ (____) _____

Parents name (if under 18): _____

Date of last visit to CHC: |__|__|/|__|__|/|__|__|

Medical record number: _____

Principal service site used: _____

Eligible list number: |__|__|__|__|__|__|

Complete name: _____
(First Name) (Last Name)

Date of birth: |__|__|/|__|__|/|__|__| Sex: 0 ☐ Male 1 ☐ Female

Contact address: _____

City: _____

State: |__|__| ZIP: |__|__|__|__|__|

Contact telephone numbers: (____) _____ (____) _____

Parents name (if under 18): _____

Date of last visit to CHC: |__|__|/|__|__|/|__|__|

Medical record number: _____

Principal service site used: _____

APPENDIX 4

BPHC CHC USER SURVEY NONRESPONSE ANALYSIS

BPHC CHC USER SURVEY
NONRESPONSE ANALYSIS
(Known **Ineligibles Excluded)**

	Number and (Percent)				
Grantee	Responded	Refused	Not Locatable	Other Nonresuonse	Total
1	26 (54.2)	1 (2.1)	20 (41.7)	1 (2.1)	48 (100.0)
2	11 (26.8)	2 (4.9)	28 (68.3)	0 (0)	41 (100.0)
3	30 (63.8)	1 (2.1)	12 (25.5)	4 (8.5)	47 (100.0)
4	49 (74.2)	1 (1.5)	15 (22.7)	1 (1.5)	66 (100.0)
5	41 (80.4)	2 (3.9)	7 (13.7)	(210)	51 (100.0)
6	46 (79.3)	3 (5.2)	8 (13.8)	(117)	58 (100.0)
7	21 (48.8)	1 (2.3)	20 (46.5)	1 (2.3)	43 (100.0)
8	30 (57.7)	3 (5.8)	15 (28.8)	4 (7.7)	52 (100.0)
9	42 (80.8)	1 (1.9)	8 (15.4)	1 (1.9)	52 (100.0)
10	39 (86.7)	2 (4.4)	4 (8.9)	0 (0.0)	45 (100.0)
11	39 (92.9)	1 (2.4)	2 (4.8)	0 (0.0)	42 (100.0)
12	38 (60.3)	7 (11.1)	15 (23.8)	3 (4.8)	63 (100.0)
13	36 (57.1)	12 (19.0)	8 (12.7)	7 (11.1)	63 (100.0)
14	38 (79.2)	2 (4.2)	6 (12.5)	(422)	48 (100.0)
15	50 (83.3)	8 (13.3)	0 (0.0)	(323)	60 (100.0)

(continued)

Grantee	Number and (Percent)				
	Responded	Refused	Not Locatable	Other Nonresponse	Total
16	35 (70.0)	7 (14.0)	(168.0)	0 (0.0)	50 (100.0)
17	25 (64.1)	0 (0.0)	12 (30.8)	2 (5.1)	39 (100.0)
18	(63840)	4 (8.0)	10 (20.0)	2 (4.0)	50 (100.0)
19	34 (85.0)	2 (5.0)	1 (2.5)	(735)	40 (100.0)
20	44 (69.8)	1 (1.6)	17 (27.0)	(116)	63 (100.0)
21	34 (75.6)	0 (0.0)	10 (22.2)	1 (2.2)	45 (100.0)
22	37 (80.4)	1 (2.2)	(178.4)	0 (0.0	46 (100.0)
23	27 (73.0)	1 (2.7)	5 (13.5)	(A)	37 (100.0)
24	42 (62.7)	4 (6.0)	18 (26.9)	3 (4.5)	67 (100.0)
25	21 (44.7)	(423)	20 (42.6)	4 (8.5)	47 (100.0)
26	47 (77.0)	5 (8.2)	5 (8.2)	4 (6.6)	61 (100.0)
27	53 (73.6)	0 (0.0)	14 (19.4)	5 (6.9)	72 (100.0)
28	47 (87.0)	1 (1.8)	(744)	(327)	54 (100.0)
29	39 (67.2)	(532)	3 (5.2)	13 (22.4)	58 (100.0)
30	33 (52.4)	5 (7.9)	14 (22.2)	11 (17.5)	63 (100.0)
31	47 (88.7)	0 (0.0)	5 (9.4)	1 (1.9)	53 (100.0)
32	41 (66.1)	5 (8.1)	12 (19.3)	4 (6.5)	62 (100.0)

(continued)

Grantee	Number and (Percent)				
	Responded	Refused	Not Locatable	Other Nonresponse	Total
33	56 (90.3)	(116)	2 (3.2)	3 (4.8)	62 (100.0)
34	42 (87.5)	0 (0.0)	5 (10.4)	(211)	48 (100.0)
35	39 (97.5)	0 (0.0)	1 (2.5)	0 (0.0)	40 (100.0)
36	42 (87.5)	3 (6.2)	1 (2.1)	2 (4.2)	48 (100.0)
37	33 (76.7)	(730)	7 (16.3)	0 (0.0)	43 (100.0)
38	51 (92.7)	2 (3.6)	2 (3.6)	0 (0.0)	55 (100.0)
39	44 (80.0)	2 (3.6)	5 (9.1)	4 (7.3)	55 (100.0)
40	60 (93.7)	0 (0.0)	1 (1.6)	3 (4.7)	64 (100.0)
41	52 (86.7)	3 (5.0)	4 (6.7)	1 (1.7)	60 (100.0)
42	52 (85.2)	1 (1.6)	4 (6.6)	4 (6.6)	61 (100.0)
43	40 (83.3)	5 (10.4)	2 (4.2)	1 (2.1)	48 (100.0)
44	50 (78.1)	5 (7.8)	9 (14.1)	0 (0.0)	64 (100.0)
45	64 (97.0)	0 (0.0)	2 (3.0)	0 (0.0)	66 (100.0)
46	53 (84.1)	2 (3.2)	2 (3.2)	6 (9.5)	63 (100.0)
47	31 (79.5)	2 (5.1)	3 (7.7)	(737)	39 (100.0)
48	47 (77.0)	(646)	8 (13.1)	2 (3.3)	61 (100.0)
Totals	**1932**	**121**	**392**	**118**	**2563**

APPENDIX 5

MEMO: DOCUMENTATION OF COMMUNITY HEALTH CENTER (CHC) USER SAMPLE SELECTION

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service
Centers for Disease Control and Prevention

National Center for Health Statistics
6525 Belcrest Road
Hyattsville, Maryland 20782

November 21, 1994

NOTE FOR Trish Royston, Jerri Regan, Joanne Lukomnik, Brenda Cox, Barbara Carlson

FROM Chris Moriarity, NCHS CM

SUBJECT Documentation of Community Health Center (CHC) User Sample Selection

The purpose of this note is to document the sample selection process for the record,

I have selected a sample of SO CHCs from a total of 10 sampling strata. Note that I have defined one additional stratum that was not present in the list in my note dated October 17 titled "CHC User Sample". The additional stratum is described below.

A listing Of the 10 sampling strata follows. It is in the same format as the list dated October 17, *in* case you would like to compare the two lists. There are some differences in the two lists, described in more detail below. As before, the stratifying variables are Census Region (1 = Northeast, 2 = South, 3 = Midwest, and 4 = West) and urban versus rural (U = urban, R = rural), as defined in the HRSA database. The rural south was large enough that I decided to split it into two pieces; UR = "R" denotes the South Atlantic Census Division, UR = "S" denotes the remainder of the South Census Region. The variable "_FREQ_" is the number of grantees in each stratum, and "T_MUSER" shows an updated count of total medical users, based on data provided to me within the last two weeks by HRSA. "MEAN" and "STD" are the average number of total medical users and the standard deviation, respectively, per grantee in each stratum. "NUMSAMP1" shows the number of sample grantees *to* be selected from each stratum under probability proportional to size sampling (pps sampling or, more technically, πps sampling, since the sampling is being done without replacement), where the measure of size is the total number of medical users. "NUMSAMP2" shows the number of sample grantees to be selected from each stratum if the allocation is proportional to the number of grantees. "SAMPINT1" and "SAMPINT2" show the sampling intervals for systematic sampling within the strata; "SAMPINT2" assumes equal probability selection of grantees with each stratum.

Note that the new stratum is listed first, indexed by REGION=0 and UR="A". This is a stratum defined exclusively by the proportion of total medical users that are managed care users. Joanne Lukomnik has consistently indicated that making estimates related to managed care users is a high priority of this survey. Recent simulation work that I did indicated that estimates related to managed care users could have high variability, given geographic stratification and pps sampling, I examined the managed care user data that I have and decided to define a stratum to contain grantees where greater than 35 percent of the

total medical users were managed care users. This was the highest proportion I could use that would still allow more than half of the managed care users in the new stratum, (The stratum would be more homogeneous for managed care users by increasing the proportion, but I felt that a "managed care user" stratum should contain at Least half of the managed care users.) Of the 18 grantees assigned to the managed care user stratum, 14 came from urban strata (5 from the Northeast, 2 from the South, 3 from the Midwest, and 4 from the West) and 4 from rural strata (one from the South Atlantic Division, 3 from the West). This past week, I did a simulation using this stratum and noticed that estimates related to managed care users became more reliable, while estimates related to race/ethnicity groups and prenatal care users became slightly less reliable.

Ideally, I should have examined various definitions of the managed care user stratum and the *corresponding* impact on estimates related to both managed care user8 and other groups. Lack of time precludes such work. Also, the data I am working with have some limitations, which impose limitation8 on the accuracy of any comparison of various stratum definitions. For example, there are several inconsistencies between the managed care *user* data and the total medical user data (more managed care users than total medical users). My understanding is that the number of managed care users is likely to increase, and the data that I have may not include recent changes. Given lack of time and data limitations, I recommended to Trish last week that we adopt the managed care stratum as defined above - this should stabilize the number of managed care users we can expect in sample, giving a corresponding stability to estimates related to managed care users.

REGION	URB	FREQ	TMUSER	MEAN	STD	NUMSAMP1	NUMSAMP2	SAMPIN1	SAMPIN2
0	A	18.	270415	15023.06	9067.13	3	2	90138.33	9.0000
1	R	32	275944	8623.25	6204.00	3	3	91981.33	10.6667
1	U	63	876964	13920.06	13304.12	8	6	109620.50	10.5000
2	R	91	666829	7327.79	5728.36	6	9	311138.17	10.1113
2	S	70	548562	**7836.60**	5124.14	**5**	7	109712.40	**10.0000**
2	u	54	734469	13601.28	9993.30	7	6	104924.14	9.0000
3	R	41	321482	7841.02	6376.04	3	**4**	107160.67	10.2500
3	u	43	514346	11961.53	11224.81	**5**	4	102869.20	10.7500
4	R	55	516327	9387.76	8401.60	5	6	103265.40	9.1667
4	U	**31**	513358	16559.94	19664.50	5	3	102671.60	10.3333
		498	5238696			**50**	**50**		

There are several other differences between the October 17 listing and today's listing. One is a change in the number of total medical users. This reflects changes in the total medical user data received from HRSA within the last two weeks. Another is a change in the sample frame from 502 grantees to 498 grantees. HRSA has determined recently that 4 grantees previously in the sample frame are out of scope for this survey.

Also, note as before that one grantee (Denver, CO, BCRR number = 080060) is being selected with certainty. This grantee is put in a certainty stratum, 1 less unit is selected *from region* 4, UR="U", and the sampling interval for that stratum is recomputed. No other grantee in that stratum has a *measure* of size that exceed8 the revised sampling interval.

Sample selection Details

The sampling process is single start pps systematic selection within strata. The presampling sort is by size (total medical users) to assure variation of size in the sample cases. The stratum samples are selected independently, One additional feature related to this sample is building in some allowance for substitutes. My assumption is that no substitution will occur unless extensive refusal conversion techniques have failed, and HRSA and I have agreed to allow the substitution. I also assume that if any eubetitution occurs, we will compute estimates both by excluding the substitutes and reweighting the "nonsubstitute" cases and by including the substitute cases, compare the two sets of estimates, look for important differences, and discus8 our strategy if we discover any important differences,

My strategy for defining substitutes is to define "clusters" of grantees within strata, followed by the random selection of one grantee from each sample cluster. I formed clusters by first sorting the grantees by size and then forming pairs. There are several problems with proceeding blindly in this faahion. First, several strata have an odd number of grantees. I decided that when I have a "leftover", I would put the 3 smallest grantees together in one Cluster. Second, there is no possible substitute for the certainty selection. Third, in several of the Strata, several grantees are so larg that pairing them together results in a "certainty" cluster being defined, One approach is to form such a cluster and treat it as a *certainty* selection. This approach struck me as somewhat artificial and clumsy, and it leads to unequal sampling weights. A second approach is to form such a cluster and allow for the possibility of multiple "hits". In the event that multiple hits occurred, no substitute would be defined. A third approach, the approach I adopted, was to not put such "large" grantees into a cluster with another grantee - any that are selected are like the certainty selection, they have no possible Substitute.

Once the clusters were formed, I performed pps sample selection of clusters (no re-eorting by size prior.03 sample selection), where the measure of size was the combined count of total medical users. Within each sample cluster, if the cluster contained more than one grantee, I randomly selected one grantee using pps sampling. Selection within each cluster was independent of selection within other clueterr. This sequence of sampling, followed by the selection of 40 users from each sampled grantee, results in equal sampling weights within a given stratum. There is variation of sampling weights across strata because the sample allocation process had to allocate an integer number of sample grantees to each stratum, and because of the certainty selection.

A listing of the sample grantees follows. The listing is sorted by stratum, and by size within stratum. WGT1 is the reciprocal of the probability of selection for the cluster containing the grantee. W WGT2 is WGT1 multiplied by the reciprocal of the probability of selection of the grantee within the cluster. Note that WGT1 and WGT2 are equal for BCRR numbers 020270 and 080060; the former was in a cluster by itself, and the latter is the certainty selection. WGT3 is the sampling weight, assuming the selection of 40 users from the count of total medical users used for sampling.

APPENDIX 6

ENCOUNTER SAMPLING INSTRUCTIONS

SAMPLE SELECTION FOR THE ENCOUNTER ABSTRACTION

This document will provide you with instructions for selecting a sample of community health center (CHC) encounters. The intention of the Encounter Abstraction is to study medical encounters at CHCs in 1994. This information will be gathered through medical records abstraction. We expect to collect data on an average of 60 encounters per CHC. First, we outline the necessary steps. Next, we provide some definitions. Finally, we provide more detailed instructions.

OUTLINE OF SAMPLING STEPS

The procedures for selecting the encounter sample are similar to those for the User Survey component. More detailed instructions follow, but the major steps are:

1. Read the instructions carefully and call your field representative with any questions

2. Compile the encounter list. This should be a list of all medical encounters (as defined below) at all eligible sites in the 1994 calendar year.

3. Exclude ineligible encounters and duplicates and obtain the total number of eligible medical encounters listed

4. Call your field representative to give the total number on your list

5. Re-order (sort) the list chronologically, if your system has this capability. If your system cannot re-order the list, simply tell your field representative and proceed with the next step.

6. Number list consecutively

7. Field representative will call you with the information necessary to select the sample from your list

8. Select the encounter sample

9. Provide record-locating information on the selected cases to your field representative

DEFINITIONS

CHC or ***grantee:*** The administrative entity that receives 330 funding from the Bureau of Primary Health Care in the Health Resources and Services Administration.

Site: A single location or facility. A CHC or grantee can have one or more sites.

Eligible ***site:*** Only those permanent sites receiving 330 funding ***and*** providing general medical services are eligible for this study. Sites that are 100% migrant worker or 100% homeless are ineligible. Sites that opened late in 1994 are also ineligible.

Encounter; A documented face-to-face contact between a patient and a provider who exercises independent judgment in the provision of services to the individual. A patient can have more than one encounter during a given visit to the center in one day.

Medical encounter: An encounter where a physician or midlevel practitioner (nurse practitioner, nurse-midwife, physician's assistant) is seen. Dental and other health encounters are not considered medical encounters, nor are services such as laboratory, x-ray, and prescriptions, unless a physician or midlevel practitioner was also seen. Nurse-only medical encounters, where the nurse is acting independently of other medical providers, are not considered medical encounters for this part of the study.

DETAILED INSTRUCTIONS

To obtain a representative sample of encounters from each CHC, it is critical that these sampling instructions be followed. Different CHCs may require slightly different procedures, depending on the level of automation of records. All acceptable procedures are described below.

1. Read the instructions carefully and call your field representative

After you have read the instructions, please call your field representative. Any questions you have can be addressed at that time. We understand that some of what we are asking may be difficult to carry out for your CHC. If you predict or encounter any problem in implementing these procedures, do not attempt an alternative solution on your own. Discuss them with your field representative, who may refer you to the project's sampling manager or one of the project's programmers, depending on the nature of the problem. The best sampling plan for your particular CHC will be determined.

2. Compile the encounter list (from all eligible sites)

The eligible encounter list must include **all** medical encounters at all **eligible** sites during calendar year 1994. The eligible list must be compiled from all available lists of encounters, whether they be computerized or hard copy lists (for example, encounter lists, encounter logs, or user files with encounter-level information). If possible, this compiled list should be treated like one big list, even if it is comprised of multiple sites. Be sure you include **all** medical encounters, including those with particularly confidential records (e.g., HIV patients) and those to specialty departments, where separate records may be maintained. If using a billing system to compile this list, be sure that all encounters are listed, not just those for which a third party was billed.

3. Exclude ineligible encounters and duplicates

To the extent possible, you must also be sure that the compiled list contains no more than one record per encounter, for those which may be listed more than once. *If* your ***system is not capable of excluding duplicates, simply tell*** *your field* ***representative*** *and proceed* ***with the next step.***

The only 1994 medical encounters that are ineligible are those that took place at ineligible sites and nurse-only encounters. All non-medical encounters (dental, other health encounters) are ineligible. (Note that encounters by most types of users considered to be ineligible for the user component of the study are **eligible** for the encounter component.) Remove all duplicate and ineligible encounters from the list.

4. Call your field representative and give list total

Determine an exact count of all encounters in the eligible list, with duplicates and ineligible encounters excluded. Then call your field representative. The number will be forwarded to the sampling manager, who will determine which records you are to select.

5. Re-order (sort) list according chronologically

If your system has the capability, re-order the encounter list chronologically by date of encounter (from January 1, 1994 to December 31, 1994). ***Tell your field representative whether you were able to re-order the encounter list chronologically and what order the encounter records were in originally.***

6. Number list consecutively

If your encounter list is on the computer, assign a 6-digit number to each encounter on the re-ordered eligible list, beginning with the number 000001. If your system cannot consecutively number your list, discuss alternatives with your field representative. ***Check that the last number on the encounter list matches the number reported to your field representative.***

7. **Field representative will call or fax you with the information necessary to select the sample from your list**

The selected cases will be listed by the number assigned in Step 6.

8. Select the encounter sample

Select the encounter sample by extracting the records corresponding to the numbers given to you by your field representative.

9. Provide record-locating information on the selected cases to your field representative

Produce the sample list on paper, including all sampled encounters in the order selected. The list should contain as much of the following information as is available from the files. ***If any particular item is difficult to attach to the sampled record or is simply unavailable, discuss this with your*** **field** ***representative.***

- encounter's assigned number (from Step **6** above)
- date of encounter
- patient's medical record number
- usual location of patient's record
- patient's date of birth
- patient's sex
- site at which encounter took place
- something to uniquely identify a particular encounter other than date:
  - provider identifier or name,
  - diagnosis or diagnostic code, or
  - time of encounter

Fax the list described above to your field representative as soon possible.

ENCOUNTER SAMPLE LIST

Name of CHC (grantee)	
BCRR ID#	
Person to contact in case of questions	

PLEASE PRINT ALL INFORMATION, FOR EACH OF THE SELECTED ENCOUNTERS.

Eligible list number: |___|___|___|___|___|___|

Date of encounter: |___|___|/|___|___|/|___|___|

Medical record number: ______________ Usual location of record: ______________

Date of birth: |__|__|/|__|__|/|__|__| Sex: 0 ☐ Male 1 ☐ Female

Site of encounter: ______________

Characteristic to identify encounter (if more than one per day) ______________

Eligible list number: |___|___|___|___|___|___|

Date of encounter: |___|___|/|___|___|/|___|___|

Medical record number: ______________ Usual location of record: ______________

Date of birth: |__|__|/|__|__|/|__|__| Sex: 0 ☐ Male 1 ☐ Female

Site of encounter: ______________

Characteristic to identify encounter (if more than one per day) ______________

Eligible list number: |___|___|___|___|___|___|

Date of encounter: |___|___|/|___|___|/|___|___|

Medical record number: ______________ Usual location of record: ______________

Date of birth: |__|__|/|__|__|/|__|__| Sex: 0 ☐ Male 1 ☐ Female

Site of encounter: ______________

Characteristic to identify encounter (if more than one per day) ______________

Eligible list number: |___|___|___|___|___|___|

Date of encounter: |___|___|/|___|___|/|___|___|

Medical record number: ______________ Usual location of record: ______________

Date of birth: |__|__|/|__|__|/|__|__| Sex: 0 ☐ Male 1 ☐ Female

Site of encounter: ______________

Characteristic to identify encounter (if more than one per day) ______________

Eligible list number: |___|___|___|___|___|___|

Date of encounter: | - | - | / | - | - | / | - | - |

Medical record number: ______________ Usual location of record: ______________

Date of birth: |__|__|/|__|__|/|__|__| Sex: 0 ☐ Male 1 ☐ Female

Site of encounter: ______________

Characteristic to identify encounter (if more than one per day) ______________

5737

Community Health Center User and Visit Study: Survey Methodology Report

MATHEMATICA
Policy Research, Inc.

http://www.mathinc.com

P.O. Box 2393
Princeton, NJ 08543-2393
(609) 799-3535

600 Maryland Ave. S.W., Suite 550
Washington, DC 20024-2512
(202) 484-9220

One Mifflin Place, 3rd Floor
Cambridge, MA 02138-4946
(617) 491-7900